Online examination with OPAL, ONYX and MAXIMA – opportunities and limitations

E. Dohmen, A. Lange, B. Kraus, S. Sturm, S. Odenbach

Professorship for Magnetofluiddynamics, Measuring and Automation Technology, Institute for Mechatronic Mechanical Engineering, Faculty of Mechanical Engineering, TU Dresden

Abstract

Der Transfer klassischer Präsenzklausuren in digitale Formate ist perspektivisch wohl unumgänglich, um als Teil einer zeitgemäßen und qualitativ hochwertigen Ausbildung auch zielorientierte sowie anspruchsvolle Leistungsüberprüfungen durchzuführen. Im Rahmen dieser Veröffentlichung skizzieren wir an konkreten Beispielen die Chancen und Grenzen der Testplattform ONYX in Zusammenspiel mit dem Computeralgebrasystem MAXIMA unter Nutzung des Lernmanagementsystems OPAL bei der Umsetzung zweier Online-Prüfungen mit bis zu 500 Studierenden. Interessant sind dabei insbesondere die beobachteten Verlagerungen der Arbeitsanteile hinsichtlich Konzeption, Klausurbetreuung und Korrektur, die wir vorstellen und deren Auswirkungen wir diskutieren.

The transfer of classical classroom examinations into digital formats is probably inevitable in the future in order to carry out goal-oriented and demanding performance examinations as part of a modern and high-quality education. In this publication, we outline the opportunities and limitations of the ONYX testing platform in conjunction with the computer algebra system MAXIMA using the learning management system OPAL in the implementation of two online examinations with up to 500 students. Of particular interest are the observed shifts in the proportion of work with regard to conception, examination supervision and correction, which we present and whose effects we discuss.

*Corresponding author: eike.dohmen@tu-dresden.de   This arcticle was originally submitted in German.

1. Introduction

In the summer semester 2020 as well as in the following winter semester, both semester-accompanying learning level self-evaluations for the students and the semester-end examinations were successfully implemented digitally at the Chair of Magnetofluiddynamics, Measuring and Automation Technology as part of the course "Measuring and Automation Technology".

The lecture series "Measuring and Automation Technology" is divided into summer and winter semesters and is completed with one examination each. Both examinations were previously divided into a basic part with 10 individual questions and a calculation part with 4 complex calculation questions. Experience has shown that the number of participants in this course drops from over 500 students to about 400 from the first semester to the following semester. Regardless of the conception of the examinations, the correction of the entire question part required about 1 person-month each semester and about 0.5 person-months for each of the arithmetic problems. This means that the correction of examinations requires about 6 person-months of personnel per year.

Based on the positive feedback of the past years on supporting asynchronous and individualisable learning offers such as lecture recordings, scripts, literature and feedback channels, the planning of a supplementary online mock examination in the teaching management system (LMS) OPAL [1] was already started in 2019 in order to analyse the advantages and disadvantages of an online examination on the one hand and to provide students with a comprehensive asynchronous possibility for individual knowledge verification and examination preparation on the other.

The LMS OPAL used and the test suite ONYX [2] integrated in it are technically developed by the e-learning service provider BPS Bildungsportal Sachsen GmbH (BPS). "*BPS Bildungsportal Sachsen GmbH was founded at the end of 2004 by 11 Saxon universities to support them in the introduction and sustainable use of new media in academic education and training. Its core task is the best possible promotion of the broad use of Internet-supported teaching/learning scenarios in Saxony's higher education institutions and their associated institutions*" [3]. The "E-Learning" *working group* of the State Rectors' Conference of Saxony *"supports the universities in continuing the path successfully taken with the "Saxony Education Portal" initiative to bundle and jointly develop their potential in the field of e-learning [...]*" [4]. The open-source computer algebra system MAXIMA is integrated into ONYX.

With the beginning of the Corona Pandemic in spring 2020, the online mock examination, which was intended as a supplementary offer, turned into an alternativeless necessity for a purely digital guarantee of teaching. In this publication, we would like to take a critical look at the potential of online examinations implemented with OPAL, ONYX and MAXIMA. For this purpose, we provide insights into our approach, present selected examples of our implementations, present problems encountered, workarounds as well as solutions and estimate the time required in order to finally initiate a discussion on the opportunities and limitations of online tests compared to classical face-to-face implementations with our experiences. This also leads to further questions: What accuracy is achieved in relation to the overarching teaching/learning objectives? What complexity can be mapped? How time-consuming is an online test?

In addition, we will discuss the Element [5] communication platform used (based on the open-source Matrix protocol), with which the accompanying communication was successfully implemented.

In this publication we will address the above points chronologically from audit conception to audit evaluation and then formulate a conclusion.

2. Concept development

The central requirement for the implementation of an online examination concept on the basis of an existing, classical face-to-face examination concept is, on the one hand, the precise redefinition of learning objectives, if necessary by extracting them from existing ex-

aminations, and, on the other hand, the derivation of new questions that can be implemented, adapted to the possibilities of OPAL, ONYX and MAXIMA, with a focus on these learning objectives. In addition to knowledge of possible types of questions (e. g. in the ONYX test environment) for examinations in STEM subjects and a technical background knowledge, the authors consider a profound technical knowledge (in the sense of programming knowledge) for the parameterisation of ONYX *questions*[1] by means of variables and for the implementation of mathematical questions with the computer algebra system MAXIMA to be essential.

Unfortunately, a central point of contact was still missing for the concept creation in the summer semester 2020 at the TU Dresden, which could clarify didactic, legal and technical questions in addition to organisational questions, actively accompany the creation of the examinations and/or arrange appropriate contacts.

With great commitment, the team of the "Centre for Interdisciplinary Learning and Teaching" (ZiLL) [6] was available in the following winter semester as a central contact point for these very questions and even provided accompanying support in the creation of tests. In addition, the "Digital Examination Task Force" was put together at the Faculty of Mechanical Engineering from voluntary experts who could advise colleagues with extensive experience on engineering-specific examination solutions.

3. Authors

In the course of the concept development, the question of the circle of authors for the generation of corresponding exams also arises, because subject-specific prior knowledge paired with the necessary specialised prior knowledge has - in the best case - only a very limited circle of people. A non-specialist eScout (a student with previous experience with the LMS OPAL used) supported us very successfully, but her and our expectations had to be adapted, because without the corresponding specialised knowledge, of course, no ONYX

questions can be generated independently and only adapted to a very limited extent. As a result, the eScouts' tasks were limited to additional work in the areas of creating illustrations and entering prepared *questions* into ONYX on the basis of detailed instructions.

eScouts should therefore be selected according to the concrete requirements and their qualifications, whereby a combination of very different qualifications (e. g. psychology and physics) would be particularly helpful, but could probably only be represented in reality by interdisciplinary eScout teams.

While the main hurdle in the summer semester 2020 was still the rapid implementation of an online examination concept from the existing classroom concept and the associated familiarisation and empowerment of (co-)authors, in the winter semester it was possible to build on the developed basis with very good technical knowledge, for example in parameterisation of *questions*, variable definitions, random values or MAXIMA functions.

In the winter semester 2020/2021, the energetic support of an eScout by the Media Centre of the TU Dresden was unfortunately no longer available to us, as they wanted to reach as many chairs as possible with the eScout programme. Considering the professional and technical skills imparted, we perceived this as a very unfavourable conception in the eScout project, as the effective benefit for our chair was very low overall. From our point of view, eScouts are very effective nuclei for personnel support and the dissemination of knowledge for the implementation of online content. Essential for this is a good to very good qualification of students as eScouts and, due to the training period, longer-term contracts of the eScouts to ensure a knowledge transfer as well as a real benefit for the educational institution. It would also be conceivable to empower student assistants from the chair.

4. Exam preparation

During the preparation of the exam, special features and bugs came to light in MAXIMA as

---

[1] Terms highlighted in **bold italics** indicate fixed terms or technical terms from OPAL or ONYX. The marking serves to differentiate from terms from general language usage.

Example: A complex question was implemented as an ONYX *TEST* and each sub-question (of this complex question) was implemented as an ONYX *QUESTION*.

well as in ONYX and OPAL, which we would now like to present together with our solutions or workarounds in the sense of a knowledge transfer.

When using the MAXIMA algebra system, some special features must be taken into account. For example, MAXIMA does not consider the commutative law for logical expressions, i.e. "a AND b" is not necessarily the same as "b AND a" and "c OR d" is not the same as "d OR c" either. In **questions** on Boolean algebra and logic**,** this is of course quite a problem! However, this is not a bug in the true sense of the word, because the MAXIMA user manual clearly states that "AND is not commutative, because the expressions a AND b and b AND a can have a different result due to operands that have not been evaluated.

Care must also be taken when using the imaginary number i (defined in MAXIMA as %i). The expression $1/i$ is not directly interpreted as equivalent to -i in MAXIMA, but must first be "simplified" by using the function expand(1/%i) in order to be recognised as equivalent to -i. There are more examples here based on features of MAXIMA. The power of this open source computer algebra system is enormous, but for effective and reliable use, one must first become familiar with the "technology" here.

When planning the exam, nomenclatures for indices, for example, should be discussed in detail and checked for their suitability. The formula character ai, for example, would be entered as a[i] in MAXIMA, but as a_i in the LaTeX text typesetting system. For numerical indices, such as a1, the notation a1 would also be permissible in MAXIMA. In the preview function in ONYX, both MAXIMA and LaTeX expressions are interpreted, which is why an identical preview image is produced (Fig. 1) and the input thus appears equivalent to the students. However, since MAXIMA would only interpret the notation a[i] as an index, only this answer would be evaluated as correct. When querying a variable, this behaviour can be intercepted by storing alternative solutions in ONYX, which would, however, only be more time-consuming to solve for formula comparisons. In order to be equitable to the students and not to include the teaching objective MAXIMA in the syllabus, indices can also be prevented by using

notations such as ai or AI - in any case, however, a consistent and uniform concept should be pursued here in assignments and solutions.

Fig. 1: Caution with indices. Different inputs by the students lead to seemingly identical results, which, however, are not interpreted in the same way by ONYX and MAXIMA.

And if one or all **questions** in ONYX suddenly stop working, it is advisable to check whether there may have just been a server update in which the ONYX or MAXIMA versions were updated (as happened during the examination phase of the 2020 summer semester, as a result of which some MAXIMA calculations became non-functional). It would be desirable to send all OPAL or ONYX users automated information about updates in advance.

But even real bugs sometimes lead to functional problems, e. g. **questions** of the type **match interaction** only worked with a maximum of four lines until July 2020, although you could insert more lines. Or with **questions** of the type **matrix interaction,** you could drag several elements on top of each other or lose elements in digital nirvana due to a programming error - here BPS reacted very quickly and was able to fix the bug in July with our advice. It is particularly annoying when **questions** can suddenly no longer be saved or the supposedly saved changes are not saved (Fig. 2).
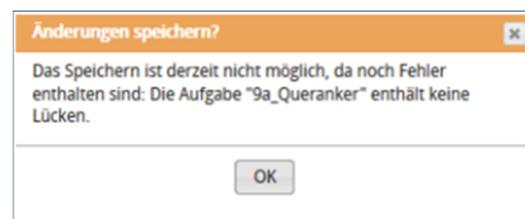

Fig. 2: Nerve-racking, because non *comprehensible ONYX errors (there were several gaps implemented in the task), which can only be circumvented by recreating the task.*

In our experience, copying the **question** helps in the best case or, unfortunately, more often only a complete recreation. And sometimes the only thing that helps with ONYX is logging

off, going to sleep and logging on, so that **questions** suddenly work overnight. We could not always clarify why some of these problems occur and due to what circumstances. According to BPS, the system is regularly cleaned up automatically at around 2:00 a.m., which would at least be an explanation for the sometimes miraculous problem solving overnight.

A frequently recurring request from our team of authors was an offline tool for test creation, from which **questions** or even entire **tests** can be exported in order to import them into the ONYX **question bank,** for example. The background to this wish is that the nested structure, cumbersome creation of several variables, long loading times for pages and instability (presumably due to high server load) are an enormous time eater and put nerves to the test.

An exam is understood as a course element **test** in OPAL. The TU Dresden has decided to create a separate OPAL instance called OPAL Exam@TUD for exams, especially for stability and data protection reasons. The course element is usually linked to an ONYX **test**, which consists of several **questions.** In addition, **sections** can be added as a structuring aid. The navigation type of a **test** to be defined as linear or non-linear has a significant effect. With a linear navigation of **questions,** students must therefore go through the exam step by step in the given pattern. A non-linear navigation would therefore be desirable, allowing students to choose an individual sequence of **questions** according to their strengths and weaknesses. However, this is not possible if only one **question** depends on another or if the solution to a previous **question** is given away with a following **question.** From the point of view of an author who is not so experienced in programming, the simple solution of defining a linear navigation within individual **sections** while being able to freely select between these **sections** in a non-linear manner seems obvious - unfortunately, this is not possible at present. In order to create a satisfactory workaround here, the MAT2 exam was therefore divided into 12 **tests** within a course element of the type **structure**, of which each **test** represents a classic complex task or a topic block. In addition, a **test** for the declaration of independence was initially inserted. In order not to

fix the time for each **test** rigidly and thus to allow an individual time allocation, the access to the course element **structure** was limited to the time for completion. However, another challenge comes into play - the consideration of different processing times for students with disadvantage compensation. Unfortunately, until the beginning of 2021, it was not possible to allocate students an additional time quota without interrupting the processing (mostly due to technical problems). For this reason, the **structure** module was duplicated in order to distinguish between students with disadvantage compensation and students with HISQIS registration (Fig. 3).
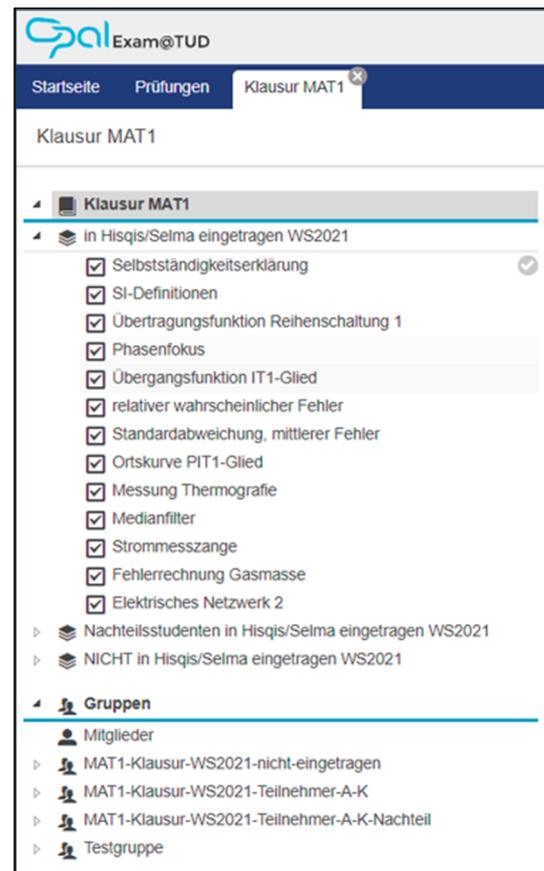
Fig. 3: OPAL "Structure" course elements as a container for a written exam with the questions as separate linear tests for combining linear and non-linear navigation.

However, the presented workaround also has its downsides, because when running the exam in exam mode, if a student has technical problems, it is first necessary to know in which test he or she was currently working in order to be able to e. g. continuation of processing.

In addition, it is necessary to repeatedly jump between tests in order to identify possible problems by means of the examination view. Furthermore, there is the aggravating side effect that when exporting the exam results, not one file but, in our case, 4 x 12, i. e. 48 individual files, have to be exported and harmonised with each other again. Consequently, it would be desirable to have an extended test structure that allows a combination of linear and non-linear contents.

In the summer semester of 2020, the work in ONYX and OPAL still surprised many special features or even bugs that were reported to BPS as the developer. In addition, wishes were expressed for the implementation of some missing functionalities. Essential points for us here were, on the one hand, the possibility to individually extend the processing time, for example, for students with disadvantage compensation (without there being any technical problems) and, on the other hand, the possibility to combine linear and non-linear test structures. This last point in particular is essential for acceptance and equivalence of online examinations. Students must have the possibility, according to their personal strengths, to individually divide up the processing time as well as the processing contents via a non-linear structure of a test. This is in contrast to the fact that more demanding, complex questions in particular can only be implemented as purely online variants with sub-questions that build on each other and have a linear navigation (without providing solutions to sub-questions on a silver platter).

It is pleasing that the individual time requirement of BPS was implemented shortly before the examination period in spring 2021. However, only in the context of one *test*. This is not possible for OPAL course components of the type *structure*, in which several *tests* are combined, see Fig. 3. Again, and therefore extremely annoying, is that the missing combination of linear and non-linear test structures affects the work of the item creators so massively! For us, this is the central weakness of the current ONYX test suite.

Unfortunately, this also resulted in the fact that the implemented individual time limit could

not be used and the known, clear additional effort was created again, because a purely linear navigation does not represent a fair examination solution in our perception. In order to be able to provide a non-linear exam that also contains linear complex questions, all questions were therefore again defined as independent *tests* within a superordinate OPAL course element *structure.* Thus, an individual time increase within an ONYX *test* unfortunately had no effect on another *test* in the same OPAL structure element, which means that the newly implemented function for our case unfortunately completely missed its actual task. Instead of an ONYX *test* with 13 *questions*, we had an OPAL course element *structure* as an exam with 13 linear ONYX *tests* as questions, as in the previous semester (cf. fig. 3). Since we also had to implement an increased processing time for students with disadvantage compensation, this exam structure had to be duplicated again in OPAL and in addition we needed another **structure** for students without HISQIS or SELMA entry, so that in the end there were a total of 39 ONYX tests.

In order to offer students supplementary security and to have a further basis for decision-making in the case of unclear errors, an ***upload interaction*** was integrated after computationally intensive sub-questions, in which students can upload handwritten notes or sketches (Fig. 4). Unfortunately, the file uploads sometimes led to significant delays under examination conditions.

Fig. 4: Questions integrated into a test for uploading handwritten notes after computationally intensive sub-questions.

5. Empowerment of students

As already explained, the syntax of MAXIMA must be learned - naturally also by the students. Therefore, it is essential to offer tests early in the semester to practise entering formulas in MAXIMA-compliant expressions, but also to practise using OPAL, ONYX, MAXIMA and feedback systems as well as your own hardware.

We also tried to alleviate this syntax problem by specifying a simple notation for variables with indices (e. g. DT = D_T), which was included in the respective **questions** as accompanying text via a global variable in the **test.** Nevertheless, questions arose several times during the exam, which is why tests accompanying the semester should not be dispensed with.

Uploading files, capturing (scanning, photographing) and compressing files or a realistic assessment of the stability of the home internet connection are also relevant.

Accordingly, uniform rules for OPAL, ONYX and MAXIMA across faculties would be desirable. Perhaps introductory courses for these tools could be offered to students with tips and suggestions for homework and online examination situations. In this way, students and teachers could build on a foundation that would be essentially identical for all courses and chairs.

6. Audit supervision

The support and supervision of students during an examination is of enormous importance in digital formats and, in our view, seems absolutely necessary for a fair and successful online examination. Questions or problems of the students can thus be clarified or possible errors in exam questions can be quickly identified. The students' feedback confirms that just being able to address problems or questions to the supervisors would give them security and have a calming effect - they would not be alone during the exam. This is a value that should not be underestimated and a positive psychological effect.

The preparation phase until the finished online exam is long and intensive - all data is revised umpteen times, checked for comprehensibility and correctness of content. Then comes the

exam and (of course) students find an error in the assignment! What to do? All students must now be informed as quickly as possible about the procedure, but how?

Fig. 5: Excerpt from an exam-related chat room: What to do if mistakes are discovered during an online exam?

Accompanying the OPALexam servers, we moderate individual chat rooms with about 80 people each at the Element instance of the TU Dresden (matrix.tu-dresden.de), which is based on the open-source messenger protocol Matrix. Over two semesters and several exams as well as internships, the platform has proven to be very reliable, fast and efficient for internal communication as well as exam supervision. The chats are predestined to distribute essential information almost instantaneously to all students. However, due to the chat structure, essential information also "wanders" quickly out of sight and must therefore be posted again regularly so as not to disadvantage anyone (Fig. 5).
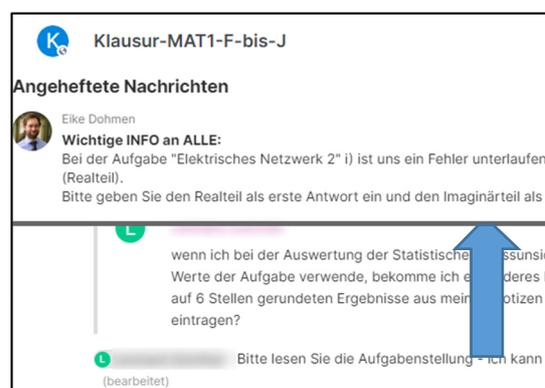

Fig. 6: Looking into the future: with the lab function "Pin", important messages can remain permanently in view for all members of a room.

During the MAT 1 exam in this winter semester, 99% of all students were successfully informed of an error in the assignment using the Matrix chat. In order to enable the permanent provision of essential information in the chat, there is currently a functionality in Matrix in the laboratory stage, whereby messages are permanently pinned directly below the room name (Fig. 6). In the future, this feature will also be available for all other users.

Protection through file upload

In addition to active support, the possibility of uploading handwritten notes or sketches, for example, is a helpful instrument to offer students additional security and to create a basis for decision-making in the case of unclear errors during correction. Accordingly, an **upload interaction** was integrated into the **tests** after computationally intensive sub-questions. Unfortunately, file uploads are a major reason for increased server load and sometimes require a relatively large amount of time, which students may then lack under examination conditions. An upload after the exam would remedy this, but it would increase the likelihood of inconsistent answers and attempted cheating.

Multi-tasking for the examiners

At the beginning of the examination period of the summer semester 2020, there were some clear performance bottlenecks in OPAL, which led to long waiting times and in one case even to a server crash during an ongoing examination, especially in the case of large examinations with a lot of "simultaneity" of user actions (such as uploads of results). Accordingly, a quick and easy solution was implemented during the exam period in summer 2020 by setting up three independent instances of OPAL exclusively for exams with the name OPALexam on servers at the TU Dresden. Unfortunately, this emergency solution remained in place for the examination period in spring 2021. As a result, the supervisors had to be logged on to three separate servers at the same time, in addition to any existing communication platforms, in order to be able to supervise the students during the exams (Fig. 7). In addition, connection problems, missing registrations or other technical problems with students could only be solved with exact knowledge of the first and last name (otherwise ambiguous), knowledge

of a possible disadvantage compensation and knowledge of the most recently completed task.
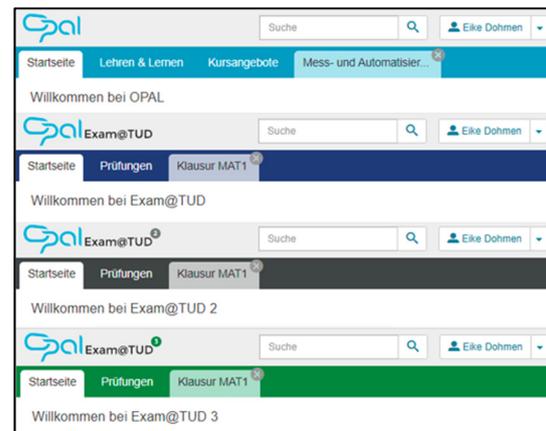

Fig. 7: OPAL, OPALexam, OPALexam2 and OPALexam3 - here multitasking of the examiners is required.

Due to these parallel structures, the supervisors had to be active simultaneously on three servers, each with three of the OPAL course elements of the type **structure** (with/without disadvantage compensation, with/without HISQIS/SELMA registration), each with 13 ONYX **tests,** and at the same time serve two communication channels (supervisors with each other, supervisors with students). If students wanted to be able to continue their work "quickly", they first had to know - as already described - the complete name, the server, the respective course element **structure** and the last **questions** worked on.

7. Examination control

Under the term **examination control,** a very helpful tool for controlling the examination and solving problems has been integrated into OPAL. Here BPS has implemented many improvements in 2020 and 2021, such as the display of the individual remaining time of students, the display of connection problems or an automatic continuation of the examination after disconnections. Nevertheless, improvements are still possible here as well. We present some selected special features or curiosities in the following.

Unfortunately, in some cases it did not seem possible to manually end the exam for a few

students who were still working on it (Fig. 8). We suspect that this is a display error and hope that BPS will clarify or remedy this.
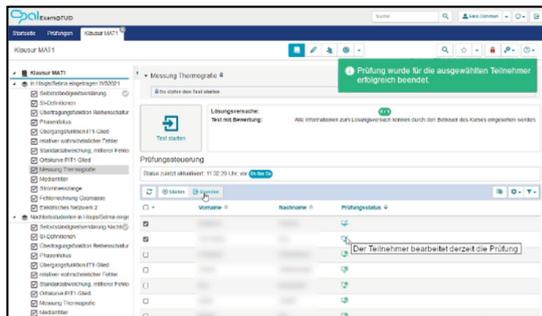

Fig. 8: Screenshot of the end of the exam - after expiry of the processing time (incl. any time buffer granted), some students continued to be displayed as "processing"; curious - even with the exam control, this status could not be changed.

A few students in our sample examination with a processing time of 1 h still had more than 1 h remaining for processing after 1 h, which should be impossible according to our understanding (Fig. 9). In these individual cases, the mock examination had to (and could) be ended manually.

Fig. 9: Screenshot of the end of the mock exam - a student still has 1h 10m remaining time for processing, although the mock exam only had 1h processing time.

Examination period variable?

A student who started the exam just before the end of processing at 10:30 am reacted very angrily to the hard, manual end of the exam at 11:15 am and posted a screenshot revealing the reason for his indignation: at SELMA (the online portal for exam administration), the exam period was incorrectly listed from 00:05 am to 23:59 pm (Fig. 10).

Fig. 10: Screenshot of a student from SELMA - the whole day is shown as the examination time window.

8. Data export

The examination team was extremely relieved when the exam was successfully completed with over 500 students without any major incidents. Traditionally, our chair then publishes the first overview statistics on the distribution of points for the question section common in presence examinations on the same evening.

The very unfavourable fragmentation of the exam over several servers and substructures unfortunately drags particularly noticeably into the phase of data export and exam correction. Now all results from 13 **tests** in 3 course modules **structure** on 3 independent servers - consequently 117 individual tests - had to be downloaded, checked and, if necessary, re-evaluated.

Just for the corresponding data archiving from OPAL including the download of the 3.8 GB ZIP archives it took (with a good internet connection) almost an hour! After that, the Excel files with the test results had to be extracted from the archives, which unfortunately was still faster than downloading the Excel files alone in OPAL (because there is only one option for displaying the PDF file, i. e. each PDF file has to be displayed first in order to be able to save it afterwards. Now one of the authors had the honourable task of copying all 117 Excel files together in order to achieve an efficient correction at least afterwards. Basically, this was relatively quick, 2-3 working hours, ...

Presets and character sets

...if it were not for the fact that on all three OPALexam instances (OPALexam, OPALexam2

and OPALexam3) the export settings would have had to be set to his personal settings BEFORE the export, in order to always receive Excel files with the appropriate character encoding (Fig. 11).

Fig. 11: System settings for export format and character set.

Specifically, three different basic settings for the export format were defined on the three OPALexam instances - apparently these basic settings are neither taken over by the existing OPAL account nor are they set uniformly. This may cause problems with data archiving (sometimes CSV, sometimes XLS) and especially with the recognition of numbers, which are interpreted as text depending on the character set!

All numbers were read in as text, which meant that all columns with numbers had to be converted into numbers (text in columns). But that was not enough, because when exporting the results of the questions with MAXIMA calculations, the ***learner response*** were exported with "." as the decimal separator, contrary to the other settings. Here are some pitfalls hidden in the data export that would vanish into thin air if only one OPAL instance with one ONYX test (for one exam) and one globally valid, simple default setting, whose effects are understandable even to the layman, were sufficient. The effort to adapt the software through BPS should therefore always be compared to the fact that many users have to wrestle with the additional expenses described.

Data loss?

With the merged table, an essential step for control and re-evaluation is made. In some cases, however, there were no variable values in the Excel table from the data archiving or these were declared as "null" (Fig. 12).

On the positive side, on the one hand the scores are correctly exported to the Excel file and on the other hand the examination view in the web interface fully documents the students' view of the examination in addition to their entries and makes it available as a PDF.

Fig. 12: Extract from the Excel evaluation table created with missing variable values (right).

A similar problem occurs when exporting floating point variables (float) from questions with MAXIMA calculations, where only 2 decimal places are exported. In an error calculation task, the variable delta_p_global_e was not exported with 1.140762589549 as specified, but ended up in Excel as 1.14. Since further calculations are made on the basis of this variable, which must be specified correctly with 6 decimal places, only an analysis in the web interface helps. This is particularly problematic if an evaluation formula was specified incorrectly and a recalculation of the results is to be carried out for all students - the corresponding input data is then missing. Since it is not yet possible to intercept such errors by the examiners through an automated re-evaluation in the web interface, each variable would have to be exported manually in order to be able to re-evaluate.

Fig. 13: Progress log as a new, powerful feature in OPAL for examination assessment

With the history log, BPS has implemented a powerful tool to track down problems based on the server logs (Fig. 13).

```
################################################################
Exam Change Log for user:                , assessmentId:        ,
course-Id: 103216254083960, coursenode-Id: 103238369591969
created 2021-02-15T10:53:36,452
################################################################
2021-02-15T10:53:36,452 INFO trainee: test started
2021-02-15T10:53:38,667 INFO exam control: status set to WORKING
2021-02-15T11:14:52,856 INFO exam control by supervisor: test finished
2021-02-15T11:14:52,866 INFO exam control by supervisor: test finished
2021-02-15T11:14:53,705 INFO exam control: status set to FINISHED
2021-02-15T11:14:56,849 INFO: test finished
```

Fig. 14: Problem analysis with the progress log - exam termination by examiner (end of exam 11:00 a.m. - termination by examiner in the superordinate OPAL course element structure for all remaining participants 11:14:52 a.m.).

With regard to the above-mentioned problem with missing variables in the export file, it was thus possible to establish that the empty cells always occurred when examiners or students cancelled an examination (Fig. 14).

```
################################################################
  Log for user:               , assessmentId:        ,
103216254083960, coursenode-Id: 103238369591969
1-02-15T10:15:44,316
################################################################
10:15:44,316 INFO trainee: test started
10:15:48,624 INFO exam control: status set to WORKING
10:47:13,661 INFO exam control: status set to WORKING, CONNECTION_LOST
10:50:45,852 INFO trainee: test resumed
10:50:48,666 INFO exam control: status set to RESUME_REQUESTED
10:53:33,547 INFO trainee: test resumed
10:53:33,665 INFO exam control: status set to RESUME_REQUESTED
10:55:07,945 INFO trainee: test resumed
10:55:08,673 INFO exam control: status set to RESUME_REQUESTED
10:56:07,251 INFO trainee: test resumed
10:56:08,685 INFO exam control: status set to RESUME_REQUESTED
10:58:14,744 INFO exam control by supervisor: test resumed, additional time: 10 minutes
10:58:18,675 INFO exam control: status set to RESUME_ALLOWED
10:59:38,681 INFO exam control: status set to RESUMED
11:10:08,706 INFO exam control: status set to FINISHED
11:10:08,899 INFO: test finished
11:14:52,910 INFO exam control by supervisor: test finished
```

Fig. 15: Problem analysis with the progress log - time expired (end of the exam with 10 minutes extra time 11:10 am).

For lines with only "none" as a variable value an expiry of the examination time during processing could be identified as the cause (Fig. 15).

9. Audit correction

In the course of redesigning our exams, we always tried to take the students' perspective into account. When correcting the exam, it then became clear that it is also necessary to look at the questions from the corrector's perspective. Which questions are efficient to correct and assess? Where do we run into decision-making problems? An example of this is the classic cloze test called ***text entry interaction*** in ONYX, in which students' vocabulary knowledge, sentence formation, dialect, existing keyboard and, last but not least, spelling must be taken into account in addition to their subject knowledge. For example, the following answers were given for a gap with the correct

answer "additive", which would have to be re-evaluated as correct or at least discussed.

- additive Y+Z=Y2	- linked
- added	- coupled
- adding	- together
- additional	- merged
- per addition	- combining
- summary	- subjunctive
- unites	- positive

This multitude of correct answers for a simple ***text entry interaction*** is hardly predictable or interceptable. With a change from a ***text entry interaction*** to ***inline choice,*** these problems could be eliminated without significantly affecting the demand or value of the task. In another task, a student informed us that he had a Swiss keyboard and therefore no "ß", which is why he used "ss" for the answers in question. This "mistake" could be caught relatively well by alternative solutions, but here, too, the use of an inline choice is a reliable alternative.
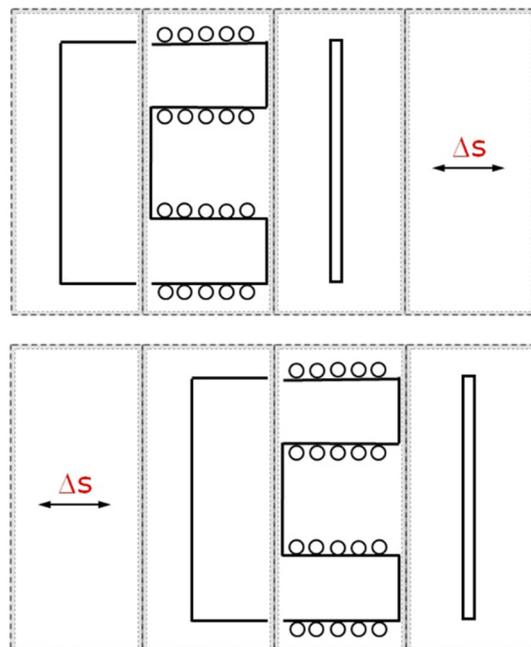

Fig. 16: Sample solution (top) and correct alternative solution (bottom) for the task: "Sketch the measuring principle of an inductive cross-armature displacement transducer".

Similarly unpredictable is the creative intelligence of the students that generates another correct solution in a supposedly clear ***match***

interaction (Fig. 16). Therefore, when assessing assignments, a manual check or correction of the answers must still take place in addition to the automated evaluation.

If a test is aborted during the processing of the last *question* (e. g. due to technical problems), it can happen that the automated assessment fails because randomly generated variables for task control are not saved. In our exam, this happened with several students in a complex task (realised via a *test* with several *questions*) for control, in which the variable {task type} was randomly assigned "Taktstrasse" or "Mischer" at the beginning of the *test.* In the case described, however, the variable {task type} was empty, although some students had already answered all questions based on this variable. The students thus received zero points on their answers, regardless of whether they were correct or incorrect. This error can only be identified by a follow-up check of the answers. The data export to an Excel file proved to be helpful here, in which incorrect solutions could be highlighted in colour using conditional formatting and complex post-assessments based on formulas were also possible. Formula-based re-assessments were particularly helpful in the case of systematic errors, e. g. IF (task type = "Taktstrasse") AND (points = 0) THEN (additional points = +0.5). Regarding the data export it should be noted that the export format (especially the character set) can be defined in the personal settings of OPAL (Fig. 11).

One drawback of the Excel export is that online post-assessments including the reasons are not exported, which is why a combined online and offline assessment is currently only possible to a limited extent. However, this issue has already been communicated to BPS and will hopefully be taken into account in the next revisions.

Online correction is advantageous for an intensive examination of each individual answer of all students, as here the correctors are given clear, concise visual feedback with all relevant information (Fig. 17). However, the response speed of OPAL for loading the pages and the nested navigation are disadvantages. Unfortunately, the visual feedback is not exported with the PDF export (e. g. with the function Archive).

The PDFs exported centrally into an archive can be opened and viewed much faster, but unfortunately this very helpful feature is missing. On the positive side, it should be noted that BPS is constantly developing the interfaces and that some new features have already been added here since the exam.

Errors happen and can only be identified through intensive, critical checks. Especially because of the complex structure of OPAL and ONYX, which is only partially transparent for the users, careless errors can hardly be avoided. Errors in content can only be identified efficiently by colleagues. However, in order to identify technical errors or problems in exams, a stochastic exam testbot (a programme that randomly generates results for a defined number of fictitious users) would be a conceivable suggestion from our point of view. This could be a programme that, for example, answers questions with random values for 100 virtual examinees, cancels tests or skips questions, etc. and then exports the resulting data as a validation data set.
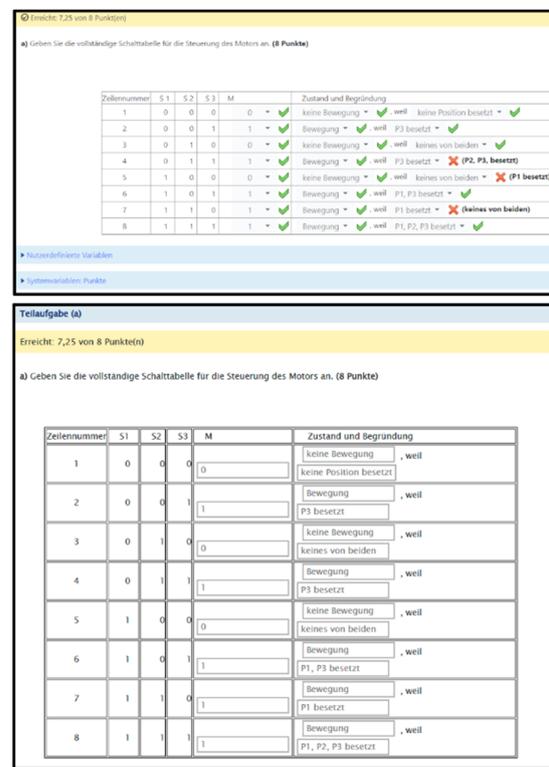

Fig. 17: Online assessment tool with visual feedback on the answers incl. the sample answer and variables (top) compared with the corresponding exported offline PDF (bottom).

Attempts to deceive

A closer look at the file names of the file uploads is interesting. On the one hand, there are indications of dubious recording sources (e. g. WhatsApp_Image_xyz) and, on the other hand, astonishing matches (e. g. "IMG20200728_081613_4.jpg"). Attempts at cheating could not be ruled out, but our perception was that very few students tried to cheat and that, especially in the summer semester 2020, the majority of students were actually rather happy to be able to take an exam at all.

In the following winter semester, however, it was claimed much more frequently that one had just mistyped or misclicked, but had evidence in the subsequently uploaded handwritten notes that one had wanted to submit something completely different. In this case, it is essential to have an unambiguous procedure that is clearly communicated before the exam in the form of framework conditions - in this specific case, that only the online entry counts and the uploads only serve as a supplement in the event of technical problems or in exceptional cases.

10. Potential assessment

In order to be able to present opportunities and limits, we would like to estimate the workload. A first (well-founded) examination statistic could already be generated a few hours after the examination. The correction of the entire question section took a total of about three person-days compared to about one person-month for the classic paper examination. Correcting the arithmetic problems took only two to three person-days per problem compared to half a person-month each. In total, the assessment effort was 12 person-days compared to the previous three person-months (approx. 60 person-days) per semester. This already includes the data export and data consolidation on a pro-rata basis. Irrespective of the enormous training and preparation effort, the pure correction time for the MAT2 exam could thus be reduced to 20%. The preparation time, on the other hand, was estimated at 6 person-months during the initial implementation and about 1-2 person-months in the following semester. This means that there was a significant additional time expenditure of up to 100% for the preparation. The cost estimate illustrates the particular potential of online examinations, especially by shifting the time spent from correction to the preparation of current questions adapted to the learning objectives, as well as making learning more flexible with increased accessibility. For teachers, the correction effort decreases dramatically, especially for exams with many participants (> 30 people). But the effort can also be worthwhile for smaller courses, since especially semester-accompanying learning status checks can be implemented in a similar way with OPAL, ONYX and MAXIMA and can then serve as the basis for an exam in a slightly modified form. In addition, the effort required to familiarise oneself with the technology and possibilities is significantly less after the first exam generation, which again reduces the time required for the preparation and implementation of exams.

OPAL & ONYX

The tools OPAL and ONYX used by the TU Dresden are developed by BPS. These tools play an important role in teaching and online examinations for our department, but also for the TU Dresden and the Free State of Saxony. On a positive note, the reliable, quick and consistently competent answers to problems, questions or suggestions should be noted. Although some of the suggestions (e. g. combination of linear and non-linear structures) are unfortunately waiting to be implemented due to the programming and associated financial effort. Here, the wishes and comments should be bundled and prioritised in a centrally coordinated manner (university-wide or statewide). Our greatest wish regarding the functionality of ONYX would be to realise a combination of linear and non-linear ONYX tests.

Framework conditions

It would be essential to define the framework conditions at the TU Dresden for online examinations in a practical and concrete way, especially in the area of conflict between examination value, data protection, identity verification, examination regulations, but also with regard to adapted further training offers for teachers and students. After one year, there is still no solution at the TU Dresden to the issue

of identity verification for online examinations that conforms to data protection regulations, which inevitably leads to a devaluation of examinations in terms of their actual informative value about the performance of an examinee and significantly reduces the value of the efforts of all those involved. It would be conceivable to create an eAssessment centre for online examinations with (several) central premises.

Central coordination office

In order to be able to offer a sustainable, competitive and attractive range of courses with online examinations across all chairs, a permanent central institution at a university with the appropriate personnel, space and technical equipment is required. This should not only be available on demand, but should proactively encourage the chairs to implement new teaching and examination concepts in the sense of excellent university development.

In this context, the "Centre for Interdisciplinary Learning and Teaching" ZiLL should be highlighted, which has taken on this task at the TU Dresden and provides competences, information, further education offers and contact persons for the implementation of online offers. It is to be hoped that the ZiLL will become an integral part of the university structure and thus be maintained in the long term.

But the commitment of the team of the TU's own messaging platform matrix.tu-dresden.de as well as the Task Force Digital Examinations, in which voluntary lecturers of the Faculty of Mechanical Engineering advise other lecturers on questions regarding the creation of examinations with their experience, should also be emphasised.

eScouts

eScouts are an excellent concept and for us represent nuclei for the transfer of knowledge about and implementation of online content. Interdisciplinary eScout teams and regular meetings of the eScouts could, in our view, help to really exploit the underlying interdisciplinary potential. At the same time, longer contracts for eScouts would make sense in order to provide financial security and ensure a transfer of know-how. It should be considered whether a corresponding financial incentive should be created for eScouts by classifying them higher in terms of salary than a student assistant - after all, further training and additional qualifications should be required to qualify as an eScout, and these should also be rewarded accordingly. Securing the basic eScout qualification, for example through an eScout Summer School or an eScout certificate, could achieve this and thus open up lucrative student jobs.

Conclusion

Regardless of the current, pandemic-related restrictions in face-to-face teaching, online examinations are thus a powerful and contemporary examination format for the authors.

11. Suggestions

In the course of the turbulent semesters, a number of ideas and suggestions have emerged with which we would like to stimulate or whose implementation would be desirable from our point of view and could lead to increases in efficiency. These are listed below in summary form:

- Combination of linear and non-linear navigation in ONYX tests
- Offline creation of ONYX tests
- Inter-faculty standards for platforms and nomenclature (MAXIMA)
- Introductory webinars in OPAL, ONYX, MAXIMA for students
- Server updates only with announcement
- a powerful OPAL instance
- Stochastic test bot
- Standardised eScout qualification
- interdisciplinary eScout teams
- Longer-term eScout employment

Acknowledgement

Literature

[1] https://www.bps-system.de/cms/products/opal-learn-management/, 20.05.2021

[2] https://www.bps-system.de/cms/produkte/onyx-testsuite/, 20.05.2021

[3] https://bildungsportal.sachsen.de/portal/parent-page/institutionen/bps-bildungsportal-sachsen-gmbh/, 20.05.2021

[4] https://bildungsportal.sachsen.de/portal/parent-page/institutionen/arbeitskreis-e-learning-der-lrk-sachsen/, 20.05.2021

[5] https://doc.matrix.tu-dresden.de/, 20.05.2021

[6] https://tu-dresden.de/tu-dresden/organisation/rek-torat/prorektor-bildung/zill, 20.05.2021