Which requirements need to be matched to allow for sound capability testing, in form of decentralized e-exams?

Introduction

The examination process is often a **tedious task** for those who are in charge. Great amounts of time go into organizational problems. Digitizing exams would resolve many of these problems. A step towards electronic examination would make the process more flexible, scalable and resource-efficient. Meanwhile, leading to a more accurate depiction of a students' competence.

Taking the current pandemic under consideration, it may not be an option to just move from paper to e-exams. Exams thus must be conducted decentralized, i.d. students take their exams at home.

It is important to notice, that *Decentralized E-exams (DE-exams)* differ from *paper based exams (PB-exams)* and even from *centralized e-exams (CE-exams)* in some key points. Foremost, the examiner has less control over the environment the exam is taken under. This raises questions about exam integrity and fairness. These questions must be addressed through careful conceptualization of questions and intelligent software design.

Digitizing exams is no new idea. Although, many concepts [cite papers that talk about byod or multi media room exams] and implementations focus on conducting e-exams in the same environment as *PB-exams*. With regard to COVID-19 this is not an option. Of course some *DE-exams* are **already conducted** today. These exams are, for the most part, making use of a **proctoring system**. In such a system a supervisor can access the examinees' device, can monitor all their activity and will watch them through their webcam. This proctoring process is costly. It hardly scales and still easily can be fooled. Further, test-taking applications are found in many *LMS' (Learn Management Systems)* such as Ilias, Moodle or Blackboard. Unfortunately, most often these applications focus on student self-assessment. They also majorly **vary in quality and utility**. As they are integrated in a complete LMS, changing to the *best* implementation is in many cases not an option. Last, as exam data is **highly confidential**, there is a strong argument to be made against closed source solutions. It is crucial to know exactly how the used application works and how data is handled. Adding, open source projects are less prone to major security issues as the development can leverage the crowdsourcing

capabilities that an open source system provides.

An overview of requirements

We find e-exams to be advatageous in a variaty of ways. Still e-exams are only valid if they can meet the same requirements that we are asked for in paperbased exams. In his book, Handke provides suitable requirements. The requirements are limited to topics on which the e-exam software has direct influence (e.g. requirements concerning exam content are not being considerd). Further these requirements can be devided into categories:

First, a requirement that defines the desired outcome of the exam:

- **General Validity**. Exams should aim to provide an accurate depiction of an examinees competence level.

Requirements that mainly influence interactions of examinees and examiners with the examination system:

- **Protection against contestation.** No formal, or technical deficiencies should accure that would question the validity of the exam.
- **Equal Treatment.** Individual examinees must be treated equally.
- **Protection against cheating.** Exams must be protected against manipulation of exam outcome by examinees.
- **Transparency.** The examination process and results must be verifiable.

Requirements that mainly influence the technical implementation of how the examination system handels data:

- **Protection of Data.** Data of examinees is persoal data, as such it must be protected from missuse.
- **Integrety.** Exam data must maintain consistency, accuracy and trustworthyness thoughout its

entire lifetime.

- **Attributability.** Taken exams must uniquely map to the examinee and vice versa.

Having identified these requirements, finding the matching design principles is the next step.

General Validity

Matching design principle: Per question time constraints to allow for partial open book exams.

As the states in their paper, examinations should support the purpose of university to produce highly capable individuales. The degree to which students succeed in that aspect is largely based on their performance in exams. Subsequentually, students are highly incentivised to focus their studies on the given exam format and questions. This interdependency between knowledge aquisition and examination shows the importancy of exam design and poses the question of what and how to test. E-exams enable assessments to more closely depict a students actual skill level by allowing the simultanieous of multiple question types.

Different question types are particularly well suited to test specific aspects of learning. These questions types can be defined as follows :

- **(Semi) Closed questions**, which mainly revolve around the demonstration of *factual knowledge*. Solutios are given by using a format like multipe choice (closed) or simple text input (semi-closed). *e.g. "What does* BYOD *stand for?"*

- **Competence questions**, which are suited to test for a certain *practical skill*. Solutions are given in form of an implementation of the specific task at hand. *e.g. "Using the software, implement an e-exam about e-learning."*

- **Essay-type questions**, which are suited for assessing *transfer knowledge* and understanding. Solutions are given by free text input. *e.g. "Explain why subjects in computer engineering are especially well suited for e-exams."*

Further different degrees of allowed aid can be identified: In open book exams, students are allowed

to solve the question at hand using any resource they feel they need. These open book exams rely mostly on both competence and essay-type questions. It could be argued that these typs of questions resemble a real world scenario in which access to information is rarely limited. Meanwhile in such an open book exam situation closed question are renderd insignificant as simple factual knowledge is easly accessible. In order to ask closed questions it is therefore necessary to restrict access to any aid.

Classical paper based exam do not provide a feasable way of combining degrees of allowed aid. Therefore, some of the question groups tend to be neglected. Resulting in constraints of possibilities to create an accurate depicture of an examinees actual competence.

With e-exams on the other hand we can implement such a varying degree of usable aid, creating a *partial* open book exam. This can be achived by letting students generally use any resource they need in order to answer the question. Additionally we introduce per question time constraints. These time constraints can be adjusted according to the question type. Leaving closed questions with a harsh time constraint, a *either you know it or you don't* situation, where there is no time to look up any solution. Essay-type questions just as competence questions can use more generous timeframes. Leaving the examinees room to make use of their tools.

Ultimatly, examination software in general has no direct impact of what exact questions the examiner asks. The content of questions obviously predifines how good a question can predict a examinees capabilities. Still, with the use of patrial open book exams, e-exams allow for a divers question set. Which allows to test factual, transfer and practical knowledge to an equally vaild degree.

Protection against contestation

Contestation of paper based exams generally is not a common problem. This is mainly due to the controlled environment paper based exams are taken under. Adding, the medium that is used to test examinees (i.e. paper) is failsave. E-exams, especially decentralized, ones introduce the possibility of failure of the exam medium. E-exams rely on software, on the operation of the exam device and of course on internet connection.

The most crutial point probably is the reliability of the e-exam software. This high reliability can

only be achived by rigorous testing and contious improvements. Of course, this applies to almost any software artefact and thus may not be considered a specific design principle per se. Further, device operatability largely lays in the responsebility of the device owner. Still, examinees should be advised to keep their devices updated, make sure they are working properly and are plugged into power.

Softwarewise these directives can be supported by making the exam accessible and workable if the internet connection is lost for short periods of time. Additionally exam answers should contiously be send to a server to minimize the possibility of data loss. In case of both a device crash and internet failure, the exam should be persistent on the local storage of the device. The device then can be rebooted and the exam can be contiued.

Equal Treatment

Equal treatment of examinees should be carried out throughout the entire examination process reaching from taking the exam to the correction of the exam.

Possible inequality arises in some key areas. In BYOD exams student devices are largely herogeneous, they run different operating systems and consist of different hardware. This fakt should not lead to different exam taking experiences. The choice of hardware should be largely irrelavant. Consequentially, it makes little sense to develop propriatary software for each operating system. Modern web technologies provide a common language among different systems. Web applications do not lack speed or functionallity and can be adopted cross platform. The software is hosted at a central instance where it can be maintained and improved and the software artefact is deliverd over a modern browser.

The process of correcting exams is also an area where possible inequalities can be found. Especially in paper based exam checking an exam for correctness is one of the most time-consuming processes in conducting an exam. Resulting in fatigue and thus sometimes falts in checking answers. Also a connection between bad hadwriting and worse grades has been found. <James 1927> shows students with bad handwriting get categorically worse grades than students with better handwriting. By using e-exams these inequalities can be fixed. First, some question types, such as multiple choice questions can be checked automatically. This is an immediate improvement over correcting these questions by hand. Second, as exam answers are available in digital text reading and checking answers is easier.

Answers must not be deciphered, correction of exams can be done faster. While also eliminating any biases against certain students.

Protection against cheating

When thinking about any assessment the consideration and handeling of academic dishonesty (e.g. cheating in an exam) is one of the most important parts. Moving from paper to e-exams poses the questions what parts – if any – must be adjusted to accomodate for changed circumstances and environments.

In his paper poses seven fields of possible cheating in exams which he then evaluates by accurance and percived severeness. Six of which are relevant for this thesisis purpose (The seventh would be *Using false excuse to delay taking test*). The fields are can be described as follows:

Student cooperation:

- **Knowing the questions** Learning what is on an exam from someone who has already taken it.
- **Cooperation with outsiders** Helping someone else cheat on an exam.
- **Cooperation with fellow examinees** Copying from another student on an exam with their knowledge.

Use of disallowed aid:

- **Exploit environmental circumstances** Copying from another student on an exam without their knowledge.
- **Use of unauthorised crib/cheat notes** Bringing prepared cheat notes to use in the exam.
- **Use of electronic, unauthorized aid** Using e.g. a smartphone to google or review lecture material.

Before thinking about how to obviate these cheating scenarios an important statement must be made: Cheating cannot completely be eliminated. There are always ways for students to engage in cheating. E-exams cannot change that. But compared to paper based exams some measures against cheating may be more effective.

Knowing a question. As the question finding process in a time consuming process, a strategy may

be to keep questions as secret as possible and reuse them throughout multiple exams. This is a rather uneffective strategie as platforms such as often provide comprehensive protocols from memory of examinees who have engaged in a given exam. E-exams can choose a different approach. Instead of having few questions and keeping them secret, e-exams can levarage large question pools. At a certain point it becomes unfeasable for students to prepare for every available question as question pools grow larger. The digital nature of these questions makes them easly shareable, allowing for larger question pools more quickly.

Cooperation with examinees. For closed questions this cooperation can be prevented by using strict time restictions. As already stated above these questions fall in the category *Either you know the answer or you don't* there is no need for a lengthy reflection period. As time constraints are tight, there really is no way of communicating with others and solving the question. For more open question types the time limitation is not as tight. At the same time answers require more in depth considerations. To ensure that students write down their own ideas and cannot share their thoughs, the copy and paste functionallity can be disabled. Further, e-exams can easily be randomized, thus preventing students form peeking or signaling solutions to eachother.

Cooperation with outsiders. As decentralized e-exams are not conducted in a controlled environment, cooperation with outsiders becomes a bigger problem. Examinees could try to take the exam in the presents of a more knowing person. Some try to solve this problem by using proctored e-exams. These exams use live survaillance through webcam and microphone that is evaluated by a person watching in realtime. This apporach hardly scales as for every 4-5 students a supervising proctor is needed. While programs like can make use of such a system, their test fees of (??250€??) leave room for addidional expenses. Although the live survaillance of stuents is not a valid option the psycological effects of being monitored can be leveraged.

Thus a measure might be to make use of integrated webcams and microphones of the devices at hand.This video and sound data can be reviewed if needed. More importantly it creates a mental barrier. If examinees really commit to engage in academic fraud they will most certaily find a way to do so. The goal is simply to prevent those from cheating that would cheat if they would feel no threat of being caught. The sole existance of any measures makes the students behave more honest. Just as

video survailance makes crime less common at public places .

Exploit environmental circumstances. Again randomization can solve this problem. As questions appear in a different order for each student even multiple choice questions can not simply be copyed.

Use of unauthorized cheat notes or electronic aid. Following the argument made about partial open book exams we find that besides time constraints no additional measures must be enforced. Cheat notes really are not of much help if there is no time to use them.

We find these cheating scenarios to be to a large degree managed by e-exams. Still, as specific software is in use, the degree of cheating must constantly assessed. Measures against bugs or security flaws must be identified.

Transparency

The examination process should be transparent for examinees. Examinees must be able to understand their mistakes and shortcomings. This of course implies that the exam software provides ways to give some kind feedback. Second, as examiners are not free of mistakes, corrections can sometimes be faulty. Well implemented transparency allows students to review the examiners correction and possibly contest against singel corrections. Important to mention is, that every student should get the chance to review their exam. The digital nature of e-exams makes this degree of transparency easly realizable. Sending a corrected digital copy of an exam, lets examinees review their answers and understand their gaps in knowledge. Contestance against singel questions could also be processed in the exam software.

Attributability, Protection of Data and Integrety

Exam data is highly sensetive and demands high levels of information security. As with any information system, basic information security principles apply. The following points prove to be of special importance.

Exam data must be uniquely tracable to examinees. This can be easly realized by having examinees log into an user account before they can perform any action. Examinees either get a unique identifier

in-software or a unique identifier that is provided by the testing authotity. Anything they do is then linked to their user id.

To assure solid data protection, strong user rights management must be enacted to assure that only authorzied groups can see or correct exams. Data in this way is largely protected from missuse. This measure ties into the intergrety of exam data. As access is restricted exam data cannot be changed. To provide even more security anserwed questions can be send to a central server instance as soon as students start answering the next question. Further frequent database backups of the exam data should be standard procedure.

Another consideration to take into account is the availabiltiy of the softwares source code. Processes should be completely transparent and comprehensible. Exam authorties should be able to host exams by them selves. This can be achived by providing the exam providing software in open source format. This has the addintional benefit that such software can benefit of crowd particpation.

Software

In the previous section I laid out areas in which e-exams can improve upon the examination pocess. As I already discussed the actual accurance of these advantages havily depends upon the software that is used in order to implement an e-exam. In the following I will take a look at common solutions:

- Ilias
- Moodle
- Blackboard
- LPlus
- OpenOlat

I will measure their quality based on their degree of fullfillment of requirements I layed out earlier.

Befor comparing these solutions there are some features that are met by every tool. Among these are:

- A way to im-/export question and to manage question pools.
- A way to implement all common question types.

- A way to automatically grade closed questions.
- A way to randomize the order of exam questions.
- A in depth documentation and/or community.
- Finished exams can be collected remotely.
- Exam data is stored at a central instance and can be accessed remotely.
- Examination is not bound to a singe location and issued over browser.

Learn Management System vs. Standalone Solution

Of the five e-exam products four are integrated into full LMS. LPlus being the exception. The degree of usability and richness, highly varies between the tools. Problematically the integration into a LMS makes them unfeasable to exchange. As LMS provide a wide range of features and applications the lock in effect a university experiences is way to large as it could justify changing to another LMS, just to improve testing capabilities. With this consideration in mind stand alone solutions have an advantage. Instead of trying to solve a whole range of e-learnig problems, a dedicated e-examination software focuses to solve a single problem. I will still compare the e-examination capabilities of big LMS as they often are the sole point of contact to e-examination of a university.

Open Source

In the context of education, large amounts of personal data are generated. Data protection and security play a big role in such an environment.

Of the considerd solutions three of them are fully open source. Blackboard only has a partially open source codebase and LPlus is a closed source software.

Timing

As I introduced the concept of partial open book exams I showed how important per question time restrictions are. Although all tools allow for time restrictions effecting the whole exam, time constraints enforceable in LPlus and OpenOlat

Connection issues

One of the biggest shortcomings of the softwares at hand is their way of handeling connection errors. All of them rely on a stable internet connection. Especially with questions that rely on time restrictions, this is problematic. Student answers can be sent to the examination authority after the time limt has expired due to delay in connectivity although a student has answered the question in time.

Test layout

As described in Section <Sth.> exams must be customizabel in such a way, that students cannot jump between questions, cannot reanswer questions and can only see one question at a the time. This customization is found in Moodle, LPlus and OpenOlat. Both Ilias and Blackboard are very limited in that regard.

Checking Identity

In decentralized exams it is not trivial to check a students identity. I disscussed possible measures in section <Sth.>. None of the tested tools allow for any identity testing beyond authentication via password. In a way this disqualifies all tools.

==========

==========

The case for e-exams

E-education is a much discussed topic. Most of the educational material has become digital. Still, paper based exams are the way to go, when it comes to assessments in German higher education. Altought some universities amonge these are the FU Berlin and the Johannes Gutenberg University Mainz have implemented some way of e-assessment, it still lacks wide application.

In its research question this thesis ask about how to imporve upon centralized e-exams. This implies, that e-exams in general are superior to paper based exams. There is good reason for this implication. It is a good idea to take a look at these reasons before moving on.

Why move from paper based to electronic exams in the first place? inherented vs. earned advantages

Put simply central e-exams replace the pen and paper of paper based exams with a computer or tablet. To give intuition of the advantage of e-exams i will focus on *bring-your-own-device* (BYOD) e-exams. In BYOD exams the device the student is taking the exam on is not provided by the universiyt authority, students bring their own tablet or laptop to take the exam on. The BYOD concept is not new. For example Robert Peregoodof talked about the BYOD implementation of the University of British Columbia in the conference.

It is important to note that moving from paper based exams to e-exams should be a strict uprgrade. Still, paper based exams and e-exams are different in some key points. Some things that were considerated best practice must be reevaluated and rethought in order to adequately fit the e-exam context. As I will show these changes do not limit the examination process but rather improve it.

When compared to paper based exams e-exams provide many advantages as disscusses in their book. I will devide these into two different types. The fist type are inheren t advantages. These are advantages that arise from soley digitizing a paper based exam. These adavantages mainly effect the efficiency of the assessment process.

The other type of advantages are created through additional considerations and by rethinking how paper based exams approach the examination process. These advantages arise mainly from improvements in testing accuracy and equality.

In order to later assess the quality of existing software solutions, after each section I will derive requirements. These requirements are needed to make use of the theoretical advantage.

Exams as a logistics problem

Thinking about the inherent advantages of e-exams, the logistical implications of exams pop into mind. To illustrate this we use the examination process at the KIT as an example . Although some steps may differ from university to university, the gist remains.

Exams must be printed and stapled. As, exam taking students are numerous, it is common for exams to take place at many different sites. Therefore, on the test day exams must be carried out to the test site. Here, the Ciw of the KIT recommends at least two supervisors per test site. On test site, exams must be distributed to students. After the exam is written, exams are collected and counted. They are then carried back to a central location, where they remain until correction. For answer checking, correctors come together, again at a central location, where they then are able to correct the exam. After correction a grade for students is published via internet. Succeeding, an exam revision for students is planned and executed. Lastly exams are archived in their paper form.

It is not supprising that handeling large amounts of paper results in logistics overhead. Removing the paper, subsequentially removes much of the logistics overhead. In an e-exam all of the exam data is digital. There is no printing, since exams are directly transferred to the students device. Further stundets answers can automatically be retrieved after the exam is over. Exam answers are digitally available. The correction of answers is no longer bound to a certain site but can be done remotely. Exam results and feedback can be directly issued to the students. Thus, revision can also be realized decentrally. The digital exam data can then be easly archived.

This comparison illustrates the advantage of e-exams with regard to logistics. Not only is there less movement of employees but more importantly there is no movement of physical paper. Further we find advantages in the digital archiving. Data can be stored space efficiantly with no need for large physical archives. Adding, archives are much saver as redundant backups are feasable and cheap.

Requirements:

- Finished exams can be collected remotely.
- Exam data is stored at a central instance and can be accessed remotely.
- Exam data can be arcived.

Statistics on the fly

Thinking of the digital nature of exam data another advantage emerges. The digital nature of e-exams allow for a fast creation of statistics. Whereas in paper exams every piece of information must be

manually digitized, e-exams are digital out of the box. Thus, analasys of exams becomes more feasible. Thinkable are statistics about general performance, but also analasys of specific questions, or student groups. As the exam is the sole indicator of a students understanding of the matter at hand, it is of utmost importance to understand where students struggle and what they are capabil of. Having easy access to exam data yields the possibility of both better exams and better courses.

Requirements:

- Possibility to create a statiscs automatically.

Questionpools

Part of the complexity and time intensiveness of exam creation lays in creating appropriate questions. Although many courses are not unique to one university, sharing of test question is not common. In paperbased exams there are no standards and there is no suitable collaboration platform. In paper based examination there is no real foundation for sharing and reusing exam questions.

E-exams have to make use of an software artefact in oder to leverage their theoretical benefits. Such a artefact allows for an enforcment of a shareable general format. If users create exams in a specific software, implementing a standard is fairly straight foreward.

> Such a standard already exits under the name of <QTI 2.2>. Having a standard allowes for educators to collaborate to create exams.

Further this collaboration must not be limited by a singel exam with only a few questions. Exams can come from question pools — a small subset of questions is selected from a way larger superset. The development and maintenance of a large question pool is very time intensive for a single person, that is why collaboration is so important. Projects in open source software show how collaboration of many can function with great success.

Large question pools do not only help in exam creation, as they allow to produce exams automatically, it also prevents students from knowing the exam beforhand. This will be further ellaborated in the section about cheating.

Requirements:

- The exam software allows for management of question pools.
- The exam software allows for export of questions, ideally for in a sharable format.

From central to decentral e-exams

In the previous section I talked about advantages of central e-exams compared to paper based exam. One key characteristic of e-exams is that they are issued via the internet. We can use this to remove any local component of e-exams, thus making them decentalized. In the following, I will talk about advantages that decentral e-exams bare and the implications of decentralizing exams on cheating and infrastructural considerations.

E-Exams as an logistics problem

As seen above, e-exams can be immensely more efficient in terms of logistics, compared to a paper based exam. There are however still some shortcomings. First, consider the allocation of students to their respective test taking location. As different students take part in different exams this allocation grows ever more complex and requires lots of planning of a central authority. Making e-exams decentral allows for this process to be streamlined enourmously.

Second, for supervision of exams human resources are needed. The for example recommends two supervisors per location. E-exams can free up some of these resources. To do this we need to take earlier considerations about partial open book exams into account. As I layed out, Partial open book exams allow for tests that do not ask for supervision, making existing supervison obsolete.

When talking about decentralizing e-exams I think it is important to say that the goal is not to force students to write their exams at home. There still can be space reserved for exams by the university. The managemant and supervision of these spaces just becomes magnitudes more easy.

Requirements:

- Examination is not bound to a singe location.

Cheating