MAKING EXAMINATIONS MORE VALID, MEANINGFUL AND MOTIVATING: THE ONLINE EXAMS SERVICE AT ETH ZURICH

Tobias Halbherr[1], Kai Reuter[2], Daniel Schneider[3], Claudia Schlienger[4], Thomas Piendl[5]

[12345]ETH Zurich, Lehrentwicklung und -technologie, Haldenbachstrasse 44, 8092 Zurich, Switzerland
[1]tobias.halbherr@let.ethz.ch, [2]kai.reuter@let.ethz.ch, [3]daniel.schneider@let.ethz.ch,
[4]claudia.schlienger@let.ethz.ch, [5]thomas.piendl@let.ethz.ch

Keywords
online exam, e-assessment, computer based assessment, summative assessment, competence orientation, authentic assessment, ILA, IT-service, learning management system, safe exam browser.

ABSTRACT

This paper provides a comprehensive overview of the institution-wide online exams service at the Swiss Federal Institute of Technology (ETH) Zurich. It is intended for stakeholders in faculty, administration, IT, and didactic support units who wish to promote and implement a similar service at their own institutions. Based on pedagogical and psychometric considerations, a case is made for how online exam environments can help to improve the quality of examinations in higher education by making performance assessments more accurate and valid, as well as more motivating and meaningful to students. Throughout the subsequent outline of the online exams service at ETH, we focus on identifying the elements critical to fulfilling these aims. This includes a discussion of the organization, development, and operation of the service and the didactic and technical support for examiners. We provide blueprints of our online exam setups, and demonstrate how even complex online exam environments incorporating a mix of applications, files and websites can be made secure. The paper finishes with a discussion and an outlook regarding our current and planned service development projects.

1. INTRODUCTION

In 2007 ETH Zurich launched the “Online Exams at ETH Zurich” project, which in 2010 culminated in an institution-wide online exams service. The service has seen continual growth from the outset, and tested approximately 4,000 students in a total of 39 online exams in Autumn Semester 2013/2014. These exams accounted for over 10% of all written examination grades in that semester. Reactions from examiners using the service are generally favorable to enthusiastic. With “online exams” we mean examinations conducted on a computer which is connected to a network (intranet or Internet) - in contrast to “offline exams”. In ETH Zurich regulations online exams are treated as a form of written examination and no special rules apply. The term “online exam” per se has no implications for an examination’s format, content or organization. Online exams may involve multiple-choice questions, essay writing, designing a machine part with CAD software, conducting a literature review using online academic databases - or anything else imaginable and technologically feasible. All online exams at ETH Zurich are organized in centrally managed computer rooms, in secure exam environments that restrict access to unwanted resources or utilities, and are always invigilated and graded. Ungraded or purely formative online tests, exercises, etc. require neither secure exam environments nor invigilation and are not referred to as “online exams” at ETH Zurich. Students in exchange programmes can participate in ETH Zurich online exams from abroad provided they organize an invigilated exam situation and a computer equipped with a secure exam client.

2. SERVICE STRATEGY

ETH Zurich's mission statement declares that "*In education, research, and services ETH Zurich measures itself against the highest recognized international standards*", and "*ETH Zurich imparts to its students the highest state of knowledge and practical skills*" (ETH Zurich, 1996). In short, ETH Zurich strives for excellence in teaching and learning. Examinations play an important role in achieving this for two main reasons. Firstly, if excellence is to be achieved, students need to meet certain minimal performance requirements. ETH Zurich, however, is not selective in admission to its (Bachelor's) degree programmes. Instead, the necessary performance requirements are ensured through examinations. In practice, this translates into fail rates in first-year examinations of up to 40% and more. It is of critical importance that such highly selective examinations measure student performance accurately and according to relevant criteria! Secondly, examinations which merely deliver proper measurements of student performance are not sufficient. In the end, the purpose of any university is to educate its students and produce highly skilled graduates. This requires motivated students who are willing to invest time and energy, engage in productive learning activities and truly cherish their subjects. Examinations should support these factors or at least not compromise them. Thus examinations also need to be adequately integrated into the greater teaching and learning context, and examination tasks need to be designed accordingly.

Based on these considerations we follow a quality-based service strategy for online exams, and have set the following goals and lead principles for the service:

Goals: (1) Online exams help to improve learning outcomes by making examinations more meaningful and motivating to students; (2) they help to improve the validity, accuracy, and relevance of performance assessments; (3) they improve the efficiency of preparing, conducting, and scoring examinations; and (4) they create additional points of contact with examiners for pedagogical support.

Lead principles: (1) Improving exam didactics and pedagogy is the driving force in the service's design. (2) Examinations are stressful, high-stakes situations, and online exam setups need to be robust, reliable, and easy to use. (3) The online exams service must meet actual customer needs: (3a) continuing service development takes place in small, manageable, and incremental steps; new exam environments are only developed when a concrete use case, a development partnership with an examiner, and a timeline towards an actual graded exam are at hand; and (3b) examiners are actively supported in exam design. (4) The service scales well, and realizes potentials for rationalizing processes, while examiners remain responsible for their exams and are not relieved of any of the duties they would normally have in traditional paper-and-pencil exams.

3. PEDAGOGY AND DIDACTICS

As mentioned above, the leading goals and principles in our service strategy first and foremost concern pedagogical and didactic considerations. In this context examinations should fulfil two requirements: they need to measure student performance accurately and according to relevant criteria, and they need to appropriately support student learning.

3.1. MEASURING STUDENT COMPETENCES

Examinations are instruments that measure student competences. To produce meaningful results examinations must fulfil several criteria: the examination results must be independent of the examiner and the examination circumstances (objectivity); they must be precise and reproducible (reliability); and they must be a valid representation of a student's competences, i.e. they must *actually* measure the competences they purport to measure (validity) (APA & NCME, 1999; Linn, 1993).

The objectivity and reliability of examinations may be compromised by a wide range of factors, such as insufficient standardization of exam procedures, exam workspaces, exam structure or exam tasks; disruptive circumstances during examinations; cognitive bias, attribution errors and vague standards in task scoring; or natural changes in scorer mood and wakefulness over time (Haladyna & Downing, 2004). Such interfering factors must be evenly distributed across examinees, limited, or eliminated completely (Halbherr & Schlienger, 2013; Coolican, 2004; Linn, 1993), and online exams offer improved potential for doing so — especially with regard to task scoring: measures to make scoring more objective and reliable, such as making students anonymous to graders, scoring on a task-by-task basis (rather than student-by-student), randomizing student sequences with every task, or the use of scoring schemes and sample solutions are much easier to integrate into the scoring workflow in online than in paper-based examinations. Moreover, in paper-based examinations one has to rely on examiners knowing of and adhering to good-practice standards, whereas once such mechanisms have been implemented in an online exam environment examiners will conform by default. Certain extraneous influences on scoring such as handwriting quality can even be eliminated altogether. Task management is facilitated by using question banks, and item statistics are readily available for quality control. In sum these advantages should amount to a considerable improvement in examination objectivity and reliability over paper-and-pencil examinations.

Where there is too much focus on objectivity and reliability in examination design the importance of examination validity is all too easily overlooked. It is tempting to take examinations at face value, assuming that if examination tasks revolve around knowledge of specific content, student performance will appropriately reflect mastery of the associated competences. Unfortunately things are not that simple. Students, for example, may have known (or remembered) a lot, yet understood nothing (Zoller & Tsaparlis, 1997). For a valid assessment of student competences the knowledge domain alone does not suffice. The cognitive processes required to solve an examination task must also be taken into account and made to reflect the cognitive processes pertaining to the competences in question (Shavelson et al., 2003). Krathwohl (2002) introduced an easy-to-comprehend revised taxonomy to hierarchically classify such cognitive processes from "lower order" to "higher order": remember, understand, apply, analyze, evaluate, and create. While closed-ended task formats such as multiple-choice questions typically score well on objectivity and reliability, designing valid closed-ended tasks for cognitive processes beyond "remember" is exacting, non-trivial and fraught with easy mistakes (Haladyna, 2004; Haladyna, Downing & Rodriguez, 2002). Furthermore, a common wish is for examinations to produce competence assessments that are valid beyond their immediate educational context, and can be generalized to performance in corresponding real-life tasks. These demands regarding examination validity are addressed in the concepts of "competence orientation" and "authentic" assessment (Gulikers, Bastiaens & Kirschner, 2004; Gielen, Dochy & Dierick, 2003; Wiggins, 1990). Designing authentic examination tasks that closely resemble corresponding real-life tasks and require examinees to directly demonstrate the respective competences is often the most dependable way for examiners to ensure validity. Online exams substantially enhance the range of task formats available to examiners by providing access to software tools, multimedia content, simulations, interactivity, web resources, or — via peripheral connections — even specialist hardware (e.g. a USB microscope), thereby greatly increasing the ability of examiners to design competence-oriented and authentic tasks in comparison to the much more limited possibilities of paper-and-pencil examinations (Hillier & Fluck, 2013). More importantly perhaps, contemporary real-life tasks are often undertaken on computers using specific software tools. Letting students do the same to solve examination tasks enhances authenticity and competence orientation by definition.

3.2. SUPPORTING STUDENT LEARNING

Examinations have long been identified as having a profound influence on student behavior and student learning (Cilliers et al., 2010; Boud, 1995; Jones, 1923). Simply put, students have a strong tendency not to learn for life, but "to the test": they will gauge their learning engagement and focus their learning activities according to the difficulty and type of examination tasks expected. A mismatch in form or content between examination tasks and learning objectives will lead to students engaging in the wrong learning activities and acquiring the wrong competences. To make matters worse, such a mismatch is likely to alienate students and erode their learning motivation. Multiple-choice tests are perhaps the most infamous example of such mismatch and the associated detrimental effects, decried for fostering surface learning and student disengagement (Struyver, Dochy & Janssens, 2005; Scouller, 1998).

For examinations to have a beneficial influence on student learning and motivation a constructive alignment of instruction, learning and assessment (ILA) is required (Biggs & Tang, 2011; Dochy & McDowell, 1998), and this triad in turn needs to relate properly to the competences defined in the learning objectives: once more, examinations should be competence-oriented and authentic (Gulikers, Bastiaens & Kirschner, 2004; Gielen, Dochy & Dierick, 2003). Competence orientation will entice students to focus their learning on the appropriate competences giving proper direction, and authenticity will let them perceive their examinations as more meaningful and compelling, reinforcing learning motivation and increasing the time and energy invested (Gulikers, Bastiaens & Kirschner, 2004; Herrington & Herrington, 1998; Sambell, McDowell, & Brown, 1997). Online exams also offer interesting new possibilities for aligning instruction, learning and assessment: for example, student exercises can easily be made available within an exam environment as a reference, and examination tasks may even build upon them directly. Last but not least, corrected examinations contain differentiated information on the respective students' learning progress, and giving students access to this information can help them to fine-tune their further learning and engage more deeply in their subjects (Higgins, Hartley & Skelton, 2002). With online exams, granting access is only a click away.

4. EXAMINATION TASKS

To assess the advantages of online exams one must first understand the benchmark against which one is comparing. Juxtaposing a multiple-choice online exam and an essay-based paper-and-pencil examination makes little sense. Instead, a multiple-choice online exam should be compared with a multiple-choice paper-and-pencil examination; task format must not be confused with examination technology. The online exams service at ETH Zurich supports computer-based alternatives for all paper-and-pencil task formats, except for freehand drawings and annotated mathematical formulae. In addition, online exams make possible new task formats that have no direct counterparts in traditional paper-based examinations, such as competence-oriented tasks involving third-party software, multimedia or web access.

4.1. TASK FORMULATION

Online exams offer enhanced possibilities over paper-based examinations in the area of task formulation: text, images, audio, video, or even interactive virtual environments may be used to define examination tasks.

4.2. CLOSED AND SEMI-CLOSED TASK FORMATS

Despite the fact that closed-ended task formats are problematic in the context of higher order cognitive processes, examination authenticity and competence orientation, they are a pervasive element in contemporary examination practice (Haladyna & Rodriguez, 2013; Frederiksen, 1984). Performing closed-ended tasks online is a real improvement over doing them on paper: scoring is flawless and immediate, examinees can change answers as often as desired without repercussions, and item statistics for quality control — perhaps more important here than in any other task format — are readily available. Closed-ended task formats offer a highly economical way to assess large groups of students objectively and reliably. The validity of closed-ended tasks is of much less concern if learning objectives indeed revolve around cognitive processes of low complexity such as remembering specifics and facts (Haladyna, 2004), though even then task performance may be influenced all too easily by cueing and educated guessing (Schuwirth, Vleuten & Donkers, 1996). Semi-closed tasks such as numerical or lexical cloze tests address this problem effectively because they cannot be solved by recognition alone, but require active recall. Purely automated scoring, however is no longer possible: 'wrong' answers must be double-checked manually to ensure that students are not scored incorrectly due to typing mistakes.

4.3. ESSAY-TYPE TASK FORMATS

Open-ended tasks can be implemented in LMSs in the form of essay questions. The examinee is given a topic, case study or asked an open-ended question and responds by writing a text. Open-ended task formats are popular and well-suited to assess the competences associated with "higher order" cognitive processes, as defined in Krathwohl's (2002) revised taxonomy (Rabinowitz & Hojat, 1989). If written by hand, however, essay tasks can be problematic for several reasons. Quality of handwriting has long been identified as having a large confounding influence on the assessment of written texts (Klein & Taub, 2005; Marshall & Powers, 1969; James, 1927) and poor handwriting legibility makes scoring cumbersome, error-prone and time-consuming. Typing is also the dominant means of composing texts today and students are usually more comfortable and familiar with typing than handwriting (Sulzenbrück et al., 2011; Mogey et al., 2010; Frand 2000). Finally, anecdotal reports from our examiners indicate a marked increase in the quality of students' task responses when switching to keyboarded texts in online exams, though research results here seem inconclusive (Cowling, 2012; Whithaus, Harrison, & Midyette, 2008; MacArthur, 2006; Russell & Plati, 2001). In sum, having students write their exam texts on the computer, rather than with paper and pencil offers substantial benefits. Indeed, we believe that essay questions currently offer the best cost-benefit ratio of all online exam task formats.

4.4. COMPETENCE-ORIENTED TASKS

While essay-type tasks are well suited to assessing the competences associated with higher order cognitive processes, they nonetheless often leave considerable room for improvement: asking an engineering student to write an essay on how he/she would design a machine part is one thing, but asking him/her to actually design the machine part is something entirely different. In the context of online exams, authenticity and competence orientation mean understanding the computer not just as a technology for exam delivery, but as a tool and working environment in its own right. Many of the competences taught at contemporary higher education institutions revolve around work processes that are defined, facilitated or enhanced by the use of IT tools (e.g. in statistics, design, programming, or simulation). Even for competences that do not directly involve IT, authenticity may still be achieved best by modelling or simulating appropriate representations of real-world tasks with the computer (e.g. in medicine, morphology or systematics, as illustrated in Figure 1).

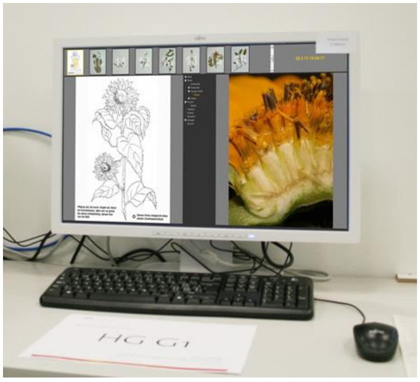

Figure 1: A virtual herbarium serves as a competence-oriented exam environment for a course in botanical systematics

4.5. OPEN- AND CLOSED-BOOK EXAMS

For some examinations competences should be highly internalized and relevant information retrievable by heart. For others, being able to find and identify relevant information on demand may be more important (Agarwal et al., 2008; Koutselini, 1997). Online exams are able to accommodate both scenarios. In online closed-book exams, all access to information beyond the formulation of exam tasks needs to be prevented. In online open-book exams access to information is typically granted via network access. Because students sitting open-book exams must be unable to delegate exam tasks to a third party, and because only certain information need generally be accessed, all means of communication must be shut down and only selective access to information should be provided. With these things attended to, open-book online exams offer a wide range of possibilities for well aligned, competence-oriented and authentic exams. The resources made available may range from course materials and course work, work templates and case studies to online repositories, academic databases and help files. In Section 5 we present technological solutions for controlling examinee access to resources during online exams.

5. TECHNOLOGY

When dealing with technology for online exams three general requirements must be kept in mind: reliability, usability, and security. Examinations are high-stakes situations, and generally both examinees and examiners are under a lot of pressure. Tolerance of technological failure is very low and the reliability of all systems involved is therefore of paramount importance. A system equipped with sufficient redundancies and (multiple) fail-safe measures is essential. Examinees will have a lot on their minds before and during an online exam, and having to figure out how an online exam environment works should not be one of them. Thus a lot of attention should be paid to good usability. The environment should not only be easy to use; examinees should also be given the chance to get acquainted with it in advance through exercises, mock exams, or formative quizzes. Finally, when sitting an online exam on an unsecured computer, fraud is all too easy, and tempting.

Online exam environments need to prevent fraud by ensuring that examinees can only access the functions and resources that are explicitly permitted. Technological security measures should not replace invigilation, however. Complete security is never possible, although a reasonable technological security benchmark is that fraud via the online exam environment should be considerably more difficult than e.g. cheating via a smartphone or a traditional cheat-sheet. Security should not compromise usability. If it does, there must be a good reason.

5.1. SAFE EXAM BROWSER (SEB)

The Safe Exam Browser (SEB) (Schneider et al., 2012; SEB website, 2014), is a free open-source software available for Windows and Mac OS X. SEB helps to impede fraud in online examinations by preventing access to unwanted resources and utilities such as system functions, websites, applications and files, and turns any computer into a secure workstation. SEB consists of two main components: (1) a kiosk application that prevents access to unwanted resources and utilities (Miller, Vandome & McBrewster, 2010); and (2) a browser to render the exam environment. The kiosk component locks down the computer by preventing access to unwanted utilities by either blocking commands (e.g. keyboard shortcuts and popup windows) or by making them inaccessible to the user by removing them from the user interface (e.g. the task manager, start menu, or third-party software). The browser component renders the actual exam environment by accessing the exam module of a learning management system via Internet or LAN.

5.2. BROWSER-BASED EXAMS WITH SEB AND LEARNING MANAGEMENT SYSTEMS (LMS)

Typical learning management systems (LMS) (McGill & Klobas, 2009) include exam or quiz modules as standard features. The LMSs Moodle and ILIAS provide specific out-of-the-box exam module extensions and configurations for secure online exams using SEB (MoodleDocs, 2014; ILIAS Feature Wiki, 2014). These extensions (1) prevent access to the online exam via any browser other than SEB; and (2) reduce the user interface of the LMS exam modules such that they only permit navigation within the exam itself and block any other undesirable LMS features such as messaging.

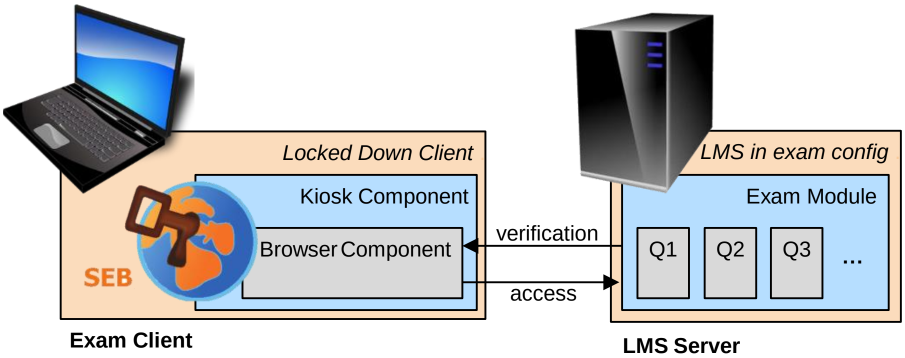

Figure 2: Schematic of the technical setup for browser-based online exams

Thus, in the browser-based exam setup the examinee is “locked into” the LMS’s exam; any additional client computer functions are being locked down by SEB; and undesired LMS functions are made unavailable by the LMS’s exam configuration. Browser-based exam setups can draw on a wide range of task formats and question types available through LMS exam modules. Figure 2 gives a schematic view of the setup.

5.3. ADVANCED EXAM ENVIRONMENTS WITH VIRTUAL DESKTOP INFRASTRUCTURE (VDI)

If additional resources and utilities — such as specific “third-party” applications (i.e. any software or application that can be run on the exam client), files or web resources — *are* desired in an online exam, SEB lockdown functions may be deactivated and access to specific applications granted. However, securing the exam client with SEB *alone* will usually no longer suffice. Contemporary applications are typically too powerful for the purposes of a secure exam. For example, most include some sort of online help browser, or provide access from within the application to the operating system's default browser. To ensure a truly sealed exam environment, such functions need to be disabled or removed. Furthermore, network and file access needs to be granted and controlled through appropriate configurations of the exam client. Such adaptations can in principle be implemented directly on physical exam clients. However, in practice this approach has proven both time-consuming and error-prone, unnecessarily complicating the development of new exam environments, as well as tedious and unreliable in operational use. A solution is required where exam clients can be swiftly and reliably reconfigured and made ready for subsequent exams (or for ordinary lab use), and where new exam environment development is independent of physical exam clients.

By implementing advanced exam environments on virtual desktops rather than physical exam clients, the problems mentioned above can be remedied successfully (Reuter, Halbherr & Piendl, 2013). The physical exam client need only install SEB together with a virtual desktop view client; thus development and testing of the exam environment is independent of the physical exam client. No tampering with the physical exam client's system is required. For each exam environment a master copy can be developed, tested and customized to fit the exact requirements of a specific exam. Before the exam, the required number of instances are copied from the master, deployed, and linked to the physical exam clients. The physical exam clients are locked down with SEB, which loads the view client for the virtual desktop instead of the browser component. SEB will also be running on the virtual desktop, in order to grant locked-down access to the LMS exam module. Examinees store their work on the virtual desktop and upload it to the LMS exam module. While the cost of setting up and maintaining a VDI is considerable, the described setup with its combination of SEB, VDI and LMS has proven very successful for implementing the more complex exam environments needed for competence-oriented examinations, thanks to its reliability, scalability, adaptability, easy customization and the comparatively low complexity and labour intensity of service maintenance. Figure 3 provides a schematic view of the setup.

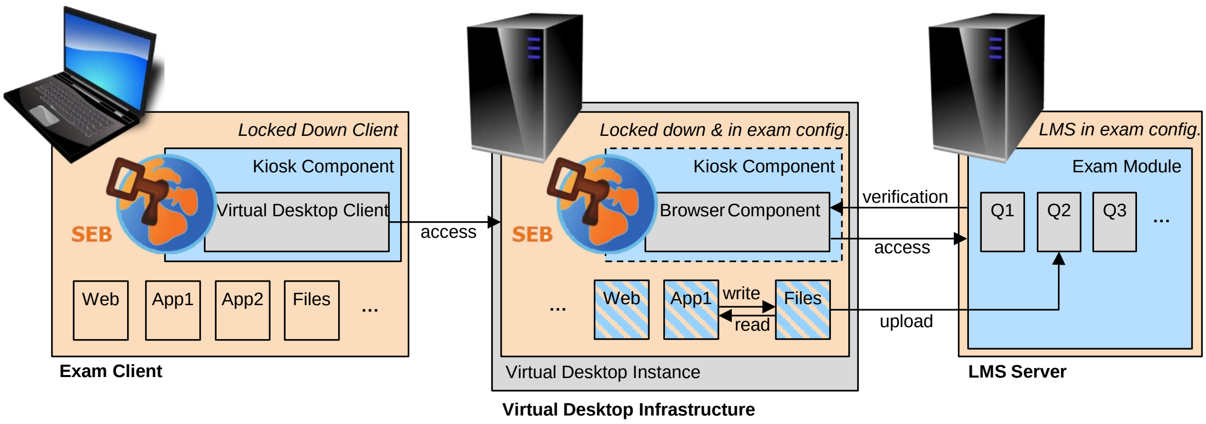

Figure 3: Schematic of the technical setup for online exam environments with virtual desktops

5.4. SCREEN RECORDING

At ETH Zurich, screen recording software is used in both browser-based exams and exams with VDI (Reuter, Halbherr & Piendl, 2013). This software takes screenshots whenever it detects a user interaction, at intervals no shorter than two seconds. It also automatically places tags and metadata in the images. These metadata, which include values on opened applications, opened files or running processes, can be searched and viewed during and after an exam. Exam sessions are recorded for three reasons. (1) The screen recording ensures transparency in case of an appeal and facilitates a fair assessment of the appeal. (2) The screen recording serves as a backup of last resort in case of unforeseen technical malfunctions or user errors that severely compromise an examinee's stored work. For example, an examinee may by accident overwrite a file containing his/her work for an exam task shortly before time runs out. The screen recording provides the examiner with the option to reconstruct the examinee's work from the screen recording, *if* the latter so desires. (3) Any technical problems that occur during an exam will also be recorded, which greatly facilitates debugging.

The screen recording is explicitly *not* intended as an additional tool for exam supervision. Examinees and examiners are informed beforehand that online exam sessions will be recorded. The recordings are strictly confidential and are only reviewed when one of the three cases mentioned above applies. The storage time of the records is strictly limited and deletion is regulated accordingly. Screen recording greatly improves the overall reliability of the online exams service as a fail-safe of last resort.

5.5. CLIENT & ROOM INFRASTRUCTURE

Online exams at ETH Zurich take place in dedicated, centrally managed university computer rooms. We consider this the most convenient solution for conducting online exams. Exam clients are easily available for testing, while hardware and software configurations are known and standardized. This effectively preempts most potential technical problems. Management of exam clients and management of online exam environments should be kept as independent of each other as possible. This reduces overall system complexity, facilitates development and debugging, and greatly improves reliability. Links between client infrastructure and online exam environments need to be tested regularly because contemporary software, especially operating systems and runtimes, are never static but require constant updates and changes which may have undesired side-effects. Beyond the technical setup, ergonomic factors need to be considered. Work stations should be conducive to student performance, help prevent fraud, facilitate invigilation, and let students enter and leave the room as efficiently as possible. Between five and ten percent of work station computers should be kept free as spares. If students sit examinations in two subsequent groups, and communication between the two groups is prevented, room capacity can effectively be doubled. ETH Zurich has four computer labs available for online exams throughout the year with a total capacity of 114 examinees (131 with spares) and one temporary dedicated online examination room with a capacity of 160 (170 with spares) available only twice a year for the duration of the examination sessions (see Figure 4). This results in an effective maximum room capacity for 548 examinees.

Figure 4: The largest room for online examinations at ETH Zurich

5.6. NETWORK AND SERVER INFRASTRUCTURE

A redundant system of load balancers, frontend webservers and backend database servers ensures the reliability, stability and scalability of the server infrastructure. All servers are virtualized. A redundant load balancer redirects participants to working frontend servers. Failed frontend servers are excluded in short time. If a physical or virtual machine fails, systems remain operational with only a few seconds downtime. The network is based on gigabit Ethernet with redundancy on the core router level. Wifi has been discarded as a network option due to its insufficient reliability and susceptibility to sabotage. The browser-based exam environment has been load-tested for up to 400 concurrent users — well above the total number of 301 online exam workstations (including spares) and the maximum number of 274 concurrent users. The VDI exam environment has been load-tested for up to 150 concurrent users and we are in the process of scaling it to a maximum of 300 concurrent users. Network and server infrastructures are routinely tested.

6. SERVICE DESIGN AND SERVICE OPERATIONS

Even if online exam environments run reliably and with very low probability of failure, many stakeholders may gauge the likelihood of severe incidents to be higher with online exams than with traditional paper-based exams. This may be due to a tendency to overestimate the likelihood of incidents in IT-based processes and underestimate the likelihood of incidents in paper-based processes (for example, paper exams may be lost at the train station or damaged beyond repair by spilled coffee; with online exams such misfortunes are impossible). These concerns need to be addressed. Responsibilities and service levels must be clearly defined. Protocols for handling severe incidents, such as aborting an examination, should be put in place in advance; even the most reliable online exams service may at some time be confronted with an aborted examination due to technical problems.

6.1. ORGANIZATION

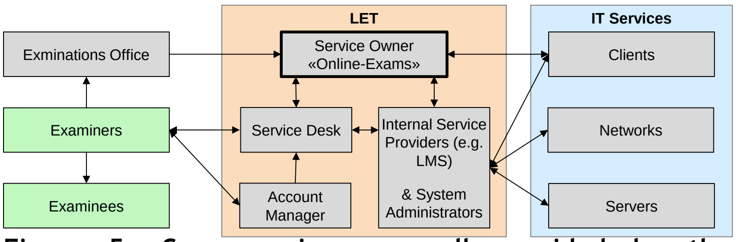

Figure 5: Core services are all provided by the Educational Development and Technology (LET) unit, allowing rapid, flexible and didactically sound accommodation of customer needs

Implementing an online exams service is an interdisciplinary, complex, time- and resource-intensive endeavour. Stakeholders include examinees, examiners and departments on the service customer and service user side, and didactic and e-learning support centers, IT services, the examinations offices, legal services, facilities, and the rectorate and executive board on the service provider side. The active support of all stakeholders is required for a well-run service; if not on board, many of the stakeholders are able to effectively stall or prevent the service from functioning entirely. We consider a clear commitment from the body superordinate to all service-providing units to be a prerequisite for a successful online exams service. This will usually be the rectorate. If superordinate support is not ensured, subordinate units may be reluctant or unable to invest adequate resources; responsibilities may remain unclear; and potential conflicts cannot be resolved properly. The service would be built on sand.

At ETH Zurich, service ownership, service design, account management, and the support desk are all located within the Educational Development and Technology (LET) unit, the university's didactics and e-learning support center. This facilitates a service strategy and service design that is didactically sound and driven by actual customer needs, rather than mere technological possibilities. Service operations are provided through a close collaboration between LET and the IT Services department. LET is responsible for the design and operation of the online exam environments, and for all customer related processes, while various IT-Services units are responsible for the client, network, and server infrastructures. Exam scheduling is organized through the Examinations Office. Coordination and communication with examinees is the responsibility of the examiners. Figure 5 gives an overview of the organization of online exam services at ETH Zurich.

6.2. PILOTING

The starting point in developing a new online exam setup at ETH-Zurich is always to pilot the setup of an examiner with a concrete use case and a timeline involving an actual graded examination. This approach has several benefits. It helps ensure that the exam environments developed meet the actual needs of both examiners and examinees. It promotes timely development, because examination dates cannot be delayed, as well as more reliable setups, because all parties involved will be more rigorous in testing when confronted with a graded exam rather than a mock exam. Because time constraints are strict, development will tend to be in smaller, manageable, incremental steps, further facilitating the development of reliable setups that meet customer demands. Finally, good usability must be ensured and examinees will always behave differently in a mock exam with no stakes in comparison to the more stressful situation of a graded exam.

A preliminary version of a setup is designed in close collaboration between LET and the examiner. In designing the exam environment, pedagogical good-practice standards, technological feasibility, and the examiner's objectives need to be taken into account. Exam environment design should be led by examination specialists, rather than IT specialists. We advise against focusing overly on technological possibilities in designing these environments. From a didactic standpoint, exam quality often benefits only marginally (if at all) from expensive technological knick-knacks, while the relevant pedagogical good-practice standards will often be easy to implement technologically. Finally, seemingly unnecessary but costly measures may be essential to guarantee exam quality.

Preliminary tests of the exam setup (with and without examinees) need to be scheduled and contingency plans prepared. If the preliminary tests produce satisfactory results the exam setup is immediately piloted (i.e. with the same group of examinees) in the graded exam; if not, the contingency plan (e.g. a paper-based version of the exam) is put into action. To further contain risks, pilots are preferably conducted with comparatively low stakes and for small-scale exams (i.e. with low failure rates and <50 examinees). After a successful pilot exam the setup usually requires additional refinement and debugging. The refined setup may then be tested in the following semester in a small number of additional small-scale graded exams.

6.3. SCALING

Transitioning a setup from a limited number of small-scale examinations to sizeable numbers of large-scale examinations scaling requires management on multiple levels: procedures, support, room infrastructure, and IT infrastructure. Here we again operate in small, manageable, incremental steps. We only scale once a setup runs smoothly, and try not to increase the respective capacities by more than double in a single scaling step. Scaling the number of examinations requires standardization of procedures and support to keep the service efficient and reliable. Measures which accommodate increasing numbers of examinations include training examiners and their assistants, creating additional redundancies in service-providing roles and staff, documentation, improved usability, and automatic process support. When scaling the number of examinees, detailed load testing of the network and server infrastructures is required. The most critical traffic loads in an online exam occur when students log on to an exam, generating hundreds of identical, quasi-simultaneous requests, and when the exams end and auto-submit. With increasing examinee numbers, the usability and reliability of the exam environment also need to increase, because support will otherwise no longer be able to cope with support requests.

6.4. SERVICE OPERATIONS

The service providers assist the examiners in designing and conducting their online exams. In this context both didactic and technological considerations are taken into account. Examiners have an account manager at their side to help them through all the required processes: planning, designing, conducting, grading and archiving their online exam. Examiners new to online exams are invited to an initial consultation session. Here the aims and goals of the online exam are discussed, an outline for its implementation is agreed upon, the further process steps are planned, and the examiners are informed on both the support available to them and their responsibilities. Examiners always remain fully in charge of their exams, and retain complete autonomy regarding exam content. LET designs and maintains the exam environments, while examiners are responsible for implementing and maintaining their online exam by using the tools provided through the service. The exam invigilators receive training in first-level technical support. It is of critical importance that any incidents that occur during an online exam be resolved quickly, and that the examinees affected may resume their work as quickly as possible. For this reason every online exam is tested beforehand, as are the exam clients, and a number of workstations are always kept free as spares. Because all student work is stored on servers in our setups, isolated incidents during an examination can always be resolved easily: if an incident cannot be resolved within one minute, the examinee is simply asked to change workstations, log back on to the exam environment, and continue the online exam. The examinee's remaining time is then extended accordingly. All the infrastructure involved in an online exam is monitored during the examination, and the responsible administrators are on standby to ensure a rapid emergency response. LET is always present with one person per examination, to assist the invigilators and to coordinate between service providers and examiners. Client management processes must be easy to administer and client configurations must be changeable in a short time, because with increasing numbers of examinations and setups there will be multiple examinations with different setups at the same time in different rooms, and/or in the same rooms in direct sequence. Student rotations need to be well organized and necessary login procedures kept clear and simple. An instruction sheet is placed at every workstation, describing the login procedure and how to handle the online exam environment. Examinees are instructed to notify support persons (i.e. the invigilators) immediately if they encounter a technical problem or error message during the online exam.

7. DISCUSSION AND OUTLOOK

Online exams help to improve the quality of assessments in higher education. They facilitate the design of valid and motivating examinations, and improve reliability and efficiency in scoring. The online exam environments in use at ETH Zurich are highly customizable and scale well, while effectively impeding fraud by securing exam clients via SEB. In service design, didactic and pedagogical considerations should be the driving force. Since examinations are high-stakes situations, and trust is easily compromised yet hard to regain, reliability should be a key concern and contingency plans need to be put in place proactively. Judging from our experience at ETH Zurich, the quality-oriented service strategy we have described pays off: the number of online exams here is (with very little advertising of our service) rapidly increasing, customer feedback is by and large satisfactory to enthusiastic, and there are even anecdotal indications that student performance and learning outcomes are improving. However, to profit from the advantages of online exams, examiners and examinees are dependent on an institution which provides such a service and the associated infrastructure.

Looking into the near future of the online exams service at ETH Zurich, we expect a continuing growth in customer demand. Thus our current main focus in service development is to improve general service scalability. We are also in the process of scaling the VDI exam environments to accommodate larger student numbers (>500), and we are preparing pilots with VDI for examinations with graphics-intensive computer aided design (CAD). An extensive review and redesign of Moodle question types is planned, as is an evaluation of online exam acceptance on the basis of regular teaching evaluations by ETH Zurich students.

8. ACKNOWLEDGEMENTS

We would like to express our special thanks to: the Managed Services team of ITS Service Delivery for their excellent cooperation; ITS System Services; ITS User Services; ITS ICT-Networks, the LET IT Services team; the SEB development team; LET Director Koni Osterwalder, ETHZ Rector Lino Guzzella and the Rectorate for their trust and support; Brigitte Schmucki for successfully launching online exams at ETH Zurich seven years ago; Sarah Frédérickx; and Katherine Hahn for English proofreading. The SEB project was conducted as part of the program "AAA/SWITCH – e-Infrastructure for e-Science" led by SWITCH, the Swiss National Research and Education Network, and was supported by funds from the ETH Board.

9. REFERENCES

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861-876.

American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. McGraw-Hill International.

Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35–48). London: Kogan Page.

Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & van der Vleuten, C. P. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in health sciences education, 15*(5), 695-715.

Coolican, H. (2004). *Research Methods and Statistics in Psychology: Fourth Edition* (pp. 54-111). London: Hodder & Stoughton Educational.

Cowling, M. (2012). Farewell to pen and paper. *Campus Review*.

Dochy, F. J. R. C., & McDowell, L. (1998). Assessment as a tool for learning. *Studies in Educational Evaluation, 23*(4), 279-298.

ETH Zurich website (1996). *Mission statement*. Retrieved February 25, 2014, from: https://www.ethz.ch/en/the-eth-zurich/portrait/self-image-and-values/mission-statement.html.

Frand, J. L. (2000). The information-age mindset changes in students and implications for higher education. *Educause review, 35*, 14-25.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*, 37-54. Dordrecht: Kluwer Academic Publishers.

Gulikers, J. T., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67-86.

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items: Third Edition* (pp. 97-126). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17-27.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education, 15*(3), 309-333.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.

Halbherr, T., & Schlienger, C. (2013). *Guidelines on Grading Written Examinations*. Zurich: ETH Zurich, Educational Development and Technology. Retrieved February 25, 2014 from: http://www.let.ethz.ch/docs/Guidelines_GradingEN_2013_11.pdf.

Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Education Research & Development, 17*(3), 305-322.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in Higher Education, 27*(1), 53-64.

Hillier, M., & Fluck, A. (2013). Arguing again for e-exams in high stakes examinations. In *2013 Australian Society for Computers in Learning and Tertiary Education Conference* (pp. 1-11).

ILIAS website (2014). *ILIAS Feature Wiki – KIOSK Modus*. Retrieved February 25, 2014, from: http://www.ilias.de/docu/goto_docu_wiki_1357_KIOSK_Modus.html?lang=en.

James, H. W. (1927). The effect of handwriting upon grading. *The English Journal, 16*(3), 180-185.

Jones, H. E. (1923). Experimental studies of college teaching: The effect of examination on the permanence of learning. *Archives of Psychology, 10*, 1–70.

Klein, J., & Taub, D. (2005). The effect of variations in handwriting and print on evaluation of student essays. *Assessing Writing, 10*(2), 134-148.

Koutselini Ioannidou, M. (1997). Testing and life-long learning: Open-book and closed-book examination in a university course. *Studies in Educational Evaluation, 23*(2), 131-139.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41*(4), 212-218.

Linn, R. L. (1993). *Educational Measurement: Third Edition* (pp. 13-146). Phoenix, AZ: The Oryx Press.

MacArthur, C. A. (2006). The effects of new technologies on writing and writing processes. *Handbook of writing research*, 248-262.

Marshall, J. C., & Powers, J. M. (1969). Writing neatness, composition errors, and essay grades. *Journal of Educational Measurement, 6*(2), 97-101.

McGill, T. J., & Klobas, J. E. (2009). A task-technology fit view of learning management system impact. *Computers & Education, 52*(2), 496-508.

Miller, F. P., Vandome, A. F. & McBrewster J. (Hrsg.) (2010). *Kiosk Software*. Saarbrücken: VDM.

Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: letting the students choose. *Research in Learning Technology, 18*(1).

Moodle website (2014). *MoodleDocs – Safe exam browser*. Retrieved February 25, 2014, from: http://docs.moodle.org/26/en/Safe_exam_browser.

Rabinowitz, H. K., & Hojat, M. (1989). A comparison of the modified essay question and multiple choice question formats: their relationship to clinical performance. *Family medicine, 21*(5), 364.

Reuter, K., Halbherr, T., & Piendl, T. (2013). *Competence-Oriented Exams using Virtual Desktop Infrastructure (VDI).* Retrieved February 25, 2014 from: http://www.switch.ch/export/sites/default/uni/projects/learn_infra/results/D4.2.7_and_7.1.3_VDI-Documentation.pdf.

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a state-mandated writing assessment. *The Teachers College Record*.

Sambell, K., McDowell, L., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349-371.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*(4), 453-472.

Schneider, D. R., Bauer, D., Volk, B., Lehre, M., Piendl, T.,: The Safe Exam Browser: Innovative Open Source Software for Online Examinations. In: *ICWE GmbH (ed.), Book of Abstracts*. Online Educa Berlin 2012 (18th International Conference on Technology Supported Learning & Training) 2012, Berlin: ICWE. ISBN 978-3-941055-16-2

Schuwirth, L. W. T., Vleuten, C. V. D., & Donkers, H. H. L. M. (1996). A closer look at cueing effects in multiple-choice questions. *Medical Education, 30*(1), 44-49.

SEB website (2014). *Concept of SEB.* Retrieved February 25, 2014, from: http://www.safeexambrowser.org/about_overview_en.html#Concept.

Shavelson, R., Ruiz-Primo, M. A., Li, M., & Ayala, C. C. (2003). *Evaluating new approaches to assessing learning*. Retrieved February 25, 2014 from: http://www.cse.ucla.edu/products/Reports/R604.pdf.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review 1. *Assessment & Evaluation in Higher Education, 30*(4), 325-341.

Sülzenbrück, S., Hegele, M., Rinkenauer, G., & Heuer, H. (2011). The death of handwriting: Secondary effects of frequent computer use on basic motor skills. *Journal of motor behavior, 43*(3), 247-251.

Whithaus, C., Harrison, S. B., & Midyette, J. (2008). Keyboarding compared with handwriting on a high-stakes writing assessment: Student choice of composing medium, raters' perceptions, and text quality. *Assessing Writing, 13*(1), 4-25.

Wiggins, G. (1990). *The Case for Authentic Assessment*. ERIC Digest.

Zoller, U., & Tsaparlis, G. (1997). Higher and lower-order cognitive skills: The case of chemistry. *Research in Science Education, 27*(1), 117-130.

10. AUTHORS' BIOGRAPHIES

Tobias Halbherr, MSc, works as an examination specialist and head of "online exams" at the Center for Educational Development and Technology (LET), ETH Zurich. He studied Psychology and Computer Science at the University of Zurich and ETH Zurich with a focus on the cognitive sciences. He graduated from the University of Zurich in 2010. His previous work includes research activities in the Visual Cognition Research Group (VICOREG) at the University of Zurich and the Center for Adaptive Security Research and Applications (CASRA), in areas of adaptive computer-supported training and assessment, human object recognition, cognitive and computational modelling, and human factors in aviation security; quantitative analyses and software development for the asset liability management and engineering treasury at Zürcher Kantonalbank; and independent work in event management, event services, and live performance video installation design.

Kai Reuter is a system administrator at the Center for Educational Development and Technology (LET), ETH Zurich. He is a technical associate for LMS systems and is responsible for the implementation and administration of VDI exam environments. Kai Reuter holds a Swiss federal VET diploma in application engineering and next to his work at LET is currently studying business informatics.

Daniel R. Schneider is project leader for the Safe Exam Browser (SEB) at the Center for Educational Development and Technology (LET), ETH Zurich. In this position, he coordinates technical aspects, public relations and contacts with other institutions who use SEB. As a software developer he is also responsible for the design, implementation and ongoing development of the Mac OS X and in part also the Windows version of SEB. Before joining ETH Zurich, Daniel studied sociology, information technology and political science, and worked in different fields (Internet media, publishing, film distribution) in the private sector as a freelancer, graphic designer, art director, software developer and journalist.

Dr. Claudia Schlienger heads the Assessment and Evaluation group at the Center for Educational Development and Technology (LET), ETH Zurich. She is responsible for teaching evaluation by students. Dr. Schlienger studied Psychology at the University of Zurich, and formerly worked in teaching and research at the Department of Informatics (University of Zurich). For her doctoral thesis she developed, deployed and evaluated a hybrid supervision model for students in a large-attendance lecture. She joined ETH Zurich in 2008 as project leader for e-collaboration.

Dr. Thomas Piendl finished his Ph.D in 1996. He holds diploma degrees in Biology and Information Science. He heads the IT-services for Teaching group at the Center for Educational Development and Technology (LET), ETH Zurich.