

GML²

2015

19.–20. März

**Grundfragen Multimedialen
Lehrens und Lernens**

E-Examinations: Chances and Challenges

Tagungsband

Nicolas Apostolopoulos, Alexander Schulz, Wolfgang Coy (Hrsg.)

Tagungsband

GML² 2015

Impressum

Herausgeber

Nicolas Apostolopoulos, Freie Universität Berlin

Alexander Schulz, Freie Universität Berlin

Wolfgang Coy, Humboldt-Universität zu Berlin

Koordination und Organisation

Irina Baskow, Freie Universität Berlin

Redaktion

Irina Baskow, Freie Universität Berlin

Sabina Pitul, Freie Universität Berlin

Alexandra Taplick, Freie Universität Berlin

Layout & Satz

CeDiS, Freie Universität Berlin

Druck

Buch- und Offsetdruckerei H. Heenemann, Berlin

Vertrieb

Waxmann Verlag GmbH

Steinfurter Straße 555, 48159 Münster

www.waxmann.com

ISBN: 978-3-8309-3331-1

Nicolas Apostolopoulos, Alexander Schulz, Wolfgang Coy (Hrsg.)

GML² 2015

E-Examinations: Chances and Challenges

Inhalt

Alexander Schulz, Nicolas Apostolopoulos, Wolfgang Coy
Vorwort zum Tagungsband GML² 2015 7

Geoffrey Crisp
Abstract: Educational advantages of computer-based examination 9

Good Practice

Robert Peregoodoff
Abstract: Large-Scale (~1400), fully online, BYOD, wireless, LMS high stakes exams:
How we do it 11

Alexander Schulz
E-Examinations in a nutshell: Lessons learnt 12

Prüfungsdidaktik

Thomas Tinnefeld
Einführung in die Prüfungsdidaktik 36

Michael Niedeggen
Abstract: Bulimielernen verhindern 50

Sebastian Schubert
Abstract: Der Progress Test Medizin: Eine multizentrische Kooperation zur Erfassung
individueller Lernverläufe im Medizinstudium 51

Sicherheit

Christoph Jeremias
Prüfungsrechtliche Rahmenbedingungen für elektronische (Präsenz-)
Prüfungen 54

<i>Kai Reuter, Tobias Halbherr</i> Kompetent und sicher: Online-Prüfungen mit Virtueller Desktop Infrastruktur und Safe Exam Browser an der ETH Zürich	63
---	----

<i>Elisabeth Katzlinger, Johann Höller</i> Verwendung der Bürgerkarte (Digitale Signatur) für E-Assessment: Realisierung von sicherheitstechnischen und rechtlichen Anforderungen an elektronische Prüfungen	75
--	----

Qualitätssicherung bei Prüfungen

<i>Heinz-Werner Wollersheim</i> Qualitätssicherung elektronischer Prüfungen mit geschlossenen Aufgaben- Formaten	89
--	----

<i>Jens Bücking</i> Qualitätsmanagement bei E-Klausuren: Risikoanalyse und QS-Maßnahmen für Strukturen, Prozesse und Inhalte	100
--	-----

<i>Susanne Bergann</i> Abstract: Wie beurteilen Studierende computergestützte Prüfungen? – Erste Ergebnisse der Evaluation der E-Examinations an der Freien Universität Berlin	113
---	-----

Prüfungssoftware

<i>Dieter Huth</i> Vergleich der gängigen Prüfungssoftwarelösungen	115
---	-----

<i>Michael Striewe</i> Automatisierte Bewertung von Programmieraufgaben: Szenarien, Werkzeuge, Perspektiven	134
---	-----

<i>Christoph Jobst</i> Potenziale neuer Fragetypen für die Naturwissenschaften	145
---	-----

<i>Marcel Dux, Oliver Tietze</i> Abstract: Mündliche Online-Prüfungen auf Distanz: Eine Diskussion der Vor- und Nachteile dieser Prüfungsform	153
---	-----

Vorwort

An deutschsprachigen Hochschulen ist spätestens seit den durch den Bologna-Prozess angestoßenen Veränderungen des Abschlusssystem von Diplom- und Magisterstudiengängen hin zu Bachelor- und Masterstudiengängen in der Mitte der 2000er Jahre das Prüfungsaufkommen massiv angewachsen. Lehrende beklagen seitdem die stark angewachsene Zahl durchzuführender Prüfungen, während Studierende wochen- bis monatelang auf die Ergebnisse der Prüfungen warten müssen. Computergestützte Prüfungen haben sich in diesem Zusammenhang in den vergangenen Jahren als ein zentrales Instrument zur zeitlichen Entlastung der Lehrenden sowie zur multimedialen Ausgestaltung von Prüfungen bewährt.

Die im Jahre 2015 als Expertenveranstaltung durchgeführte GML²-Tagung zu Grundfragen Multimedialen Lehrens und Lernens richtete sich daher an Akteure aus Lehre, Forschung und Wirtschaft, die sich mit computergestützten Prüfungen befassen. Nationale und internationale Expertinnen und Experten aus Deutschland, der Schweiz, Österreich, Kanada und Australien legten ihre Erfahrungen aus den folgenden Teilbereichen computergestützter Prüfungen dar: Good-Practice-Szenarien, Prüfungsdidaktik, juristische Perspektive und Sicherheit, Sicherung der Prüfungsqualität, technische Konzeptionen und auf dem Markt derzeit erhältliche Prüfungssoftwarelösungen.

Mit dem vorliegenden Tagungsband stellen wir im Nachgang gerne auch die Artikel und Abstracts der auf der Tagung gehaltenen Vorträge zur Verfügung. Wir möchten uns an dieser Stelle ausdrücklich bei allen Referentinnen und Referenten für ihre spannenden und wegweisenden Beiträge und bei allen Teilnehmerinnen und Teilnehmern für ihr Interesse und die äußerst engagierten Diskussionen bedanken. Zuletzt geht ein großer Dank an die Personen, die die Vorbereitung und Durchführung der Tagung und die Erstellung des Tagungsbands tatkräftig unterstützt haben.

Berlin, Oktober 2015

Keynote

Geoffrey Crisp
Royal Melbourne Institute of Technology, Australia

Educational advantages of computer-based examination

Universities need to explore the 21st century skills and capabilities that our students will need as they live and work in a world dominated by ubiquitous technology, increasing uncertainty and integrated complexity. As universities, we need to expand our repertoire of assessment tasks so they include a more sophisticated use of all learning environments, both physical and virtual. As we redesign our curricula to encompass the opportunities afforded by technology and instant access to almost unlimited information sources we will need to redesign our assessment tasks and formats so that students are able to appropriately demonstrate learning at the more sophisticated end of the SOLO or Bloom's taxonomy.

Online assessment and e-exams are two recent examples of a move to more diversity in our assessment formats. Much of the work on online assessment and e-exams has focused on the technical, security and efficiency issues as well as comparisons of students' performances on the same pedagogical task in digital and physical environments. At the GML² 2015 conference we looked at the assessment task itself and reviewed why we need to move to more online assessment and e-exams because they offer opportunities for more sophisticated assessment tasks that are more appropriately aligned to the 21st century skills and capabilities that our students will need.¹

References

Grisp, G. T. (2014). Assessment in next generation learning spaces. In K. Fraser (Hrsg.), *The future of learning and teaching in next generation learning spaces* (Bd. 12, S. 85-100). Bingley: Emerald Group Publishing Limited.

Transforming assessment. Abgerufen unter <http://transformingassessment.com/> [06.02.2015]

Transforming exams: A scalable examination platform for BYOD invigilated assessment. Abgerufen unter <http://transformingexams.com> [06.02.2015]

¹ Crisp, G. T. (2015). *Educational advantages of computer-based examination: Keynote auf der Tagung "GML² 2015: Grundfragen Multimedialen Lehrens und Lernens – E-Examinations: Chances and Challenges*." Abgerufen unter <http://www.gml-2015.de/tagungsprogramm/crisp/> [30.06.2015]

Good Practice

Robert Peregoodoff
University of British Columbia

Large-Scale (~1400), fully online, BYOD, wireless, LMS high stakes exams - How we do it

Since 2010, the Sauder School of Business has been conducting fully online assignments, midterm and final exams as large as ~1400 students concurrently. It is important to note that these exams are conducted under normal face-to-face invigilated conditions (i. e. no remote or distance exams) and that it is the responsibility of the students to provide a suitable device to sit for the exam. All exams are conducted using the University of British Columbia (UBC) Learning Management System (LMS) which is currently Blackboard Learn. From 2010 until 2012 all exams were conducted in Blackboard (WebCT) Vista. Normal duration for the exams is between 90 and 150 minutes.

Sauder has to date had over 21,000 individual student exams, resulting in significant savings to the institution as well as an improved assessment strategy owing to the ability to leverage the online medium. Normal first iterations of exams for new instructors are a straight “conversion” from the existing paper versions, however experience has been by the 2nd or 3rd iteration, faculty are now writing exam questions that would not be possible using paper and pencil.

The presentation¹ at the GML² conference covered the six primary components of the program:

- Logistics → what does it take to reduce technical failure to 0.005%?
- Design → what types of questions can be written for the online medium?
- Security and Integrity → Prevention or detection? Or both?
- Data → through a specific exam, the analytics and data available will be presented
- Technical Specifications → overview of the UBC Wireless network that is the backbone of this program
- Culture → what does it take to bring risk-adverse faculty to this program

¹ Peregoodoff, R. (2015). *Large-Scale (~1400), fully online, BYOD, wireless, LMS high stakes exams: How we do it: Vortrag auf der Tagung “GML² 2015: Grundfragen Multimedialen Lehrens und Lernens – E-Examinations: Chances and Challenges.”* Abgerufen unter <http://www.gml-2015.de/tagungsprogramm/peregoodoff/> [30.06.2015]

E-Examinations in a nutshell: Lessons learnt

Zusammenfassung

An den Hochschulen im deutschsprachigen Raum werden computergestützte Prüfungen seit einigen Jahren aus unterschiedlichen Perspektiven als viele Hoffnungen beflügelnde Technisierungsmethode für den Hochschulalltag betrachtet.

Einer der Blickwinkel auf computergestützte Prüfungen fokussiert die Lernenden und somit die didaktischen Aspekte. Lehre an den Hochschulen soll dieser Sicht zufolge mittels ganzheitlicher E-Learning-Komponenten, bei denen zu den Blended-Learning-Ansätzen eben auch computergestützte Prüfungen (E-Examinations) gehören, modernisiert werden. Ziel der didaktischen Ansätze ist, den Lernenden z. B. anhand von Fallbeispielen das kanonische Wissen anschaulicher zu vermitteln und sie in den Prüfungssituationen mit den gleichen computergestützten Werkzeugen arbeiten zu lassen, die sie auch während des Semesters schon zum Lernen eingesetzt haben.

Eine andere Perspektive konzentriert sich auf die Lehrenden. Diese haben in den letzten zehn Jahren mit stark angestiegenen Prüfungszahlen zu kämpfen. Lehrende sehen in computergestützten Prüfungen zunächst die zeitlichen Entlastungseffekte. Korrekturen und Bewertungen sollen mit computergestützten Prüfungen erheblich beschleunigt werden. Lehrende hätten somit wieder mehr Zeit, die sie in Lehre und Forschung investieren können.

Die Leitungen der Hochschulen heben bei computergestützten Prüfungen wiederum die Change-Management-Perspektive hervor. Hochschulinterne Verwaltungsabläufe sollen effizienter gestaltet werden, um einerseits Lehrende zu entlasten und andererseits Lernenden schneller als zu analogen Zeiten Rückmeldungen über ihre Prüfungsergebnisse geben zu können.

Das Center für Digitale System (CeDiS) der Freien Universität Berlin befasst sich seit 2005 mit diesen unterschiedlichen Blickwinkeln auf die Computerisierung des Prüfungswesens. Standen in den ersten Jahren zunächst stark spezialisierte didaktische Aspekte für einzelne Institute im Vordergrund, so wurden seit 2007 stärker effizienzsteigernde Aspekte, die für viele Fachbereiche anwendbar sein sollten, fokussiert.

Was explorativ am CeDiS als Spin-Off-Thema aus dem Bereich E-Learning begann, entpuppte sich alsbald als ein weites und sehr eigenes Feld, das sich mittlerweile in viele eigene Arbeitsthemen untergliedert. Dazu gehören logistische Fragen (also wie mit großen Kohorten von mehreren hundert bis einigen tausend Studierenden umzugehen sei), die Prüfungsdidaktik (die sich der Frage widmet, wie Lerninhalte sinnvoll geprüft werden können), die rechtlichen Fragen (wozu dann das Prüfungsrecht und der Datenschutz gehören), Fragen der Qualitätssicherung und Effizienzsteigerung (die durch Evaluationen und wissenschaftliche Begleitforschung empirisch fundiert werden müssen) und die Technik (mitsamt der technischen Sicherheit).

Am CeDiS wurde im Laufe der Jahre sukzessive ein eigener Arbeitsbereich E-Examinations institutionalisiert, der sich ab 2011 der Konzeption und dem Aufbau eines computergestützten Prüfungszentrums widmete: dem E-Examination Center (EEC). Im Februar 2013 wurde das EEC mit 151 Prüfungsplätzen als eines der größten computergestützten Prüfungszentren im deutschsprachigen Raum eröffnet. Im EEC wurden seit der Eröffnung bis heute (August 2015) mehr als 25.000 Studierende und Studienbewerber/innen der Freien Universität Berlin (FUB) und der Charité Universitätsmedizin Berlin (CUB) computergestützt geprüft.

Der folgende Artikel fasst zusammen, welche Erfahrungen am CeDiS seit 2005 im Bereich computergestützter Prüfungen gemacht und welche Schlüsse daraus für die weiteren Herangehensweisen an das Thema computergestützter Prüfungen gezogen wurden. Es finden sich darüber hinaus Zusammenfassungen der von der Kollegin Susanne Bergann ebenfalls auf der GML² Tagung 2015 vorgestellten und bislang nicht veröffentlichten Evaluationsergebnisse der inneruniversitären Arbeitsstelle Lehr- und Studienqualität. Der Artikel endet mit einem kurzen Ausblick auf einige Teilaspekte computergestützter Prüfungen, auf die sich CeDiS in den nächsten Jahren konzentrieren wird.

Exploration (2005–2006)

Prüfungen mit Notebooks in der Mensa und in PC-Pools

Am Anfang waren computergestützte Prüfungen an der Freien Universität stark von dem didaktischen Anspruch geleitet, die Lehre durch Blended-Learning-Szenarien ganzheitlich vervollständigen zu können. Um Medienbruch zu vermeiden, sollten die gleichen Instrumente in den Prüfungssituationen eingesetzt werden können, die auch schon während des Semesters in die Lehre eingebunden wurden. Gleichzeitig sollten Lehrende weniger Zeit mit der Bewertung der Prüfungsleistungen verbringen müssen.

Prüfungssoftware

Da bereits die Statistiklehre der Wirtschaftswissenschaft der Freien Universität als Blended-Learning-Veranstaltung umgestellt war, wurde als paradigmatischer Pilotversuch auch die Semesterabschlussprüfung der Statistik computerisiert. In den ersten computergestützten Prüfungen in den Jahren 2005 und 2006 kamen die gleichen Tools zum Einsatz, die auch in der Lehre verwendet wurden: die PC-Software *Statistiklabor* und das webbrowsers-basierte Gesamtcurriculum *Neue Statistik*. Mit dem Statistiklabor konnten Studierende in der Prüfung Datensätze fallbasiert bearbeiten, auswerten und interpretieren. Das didaktisch motivierte Statistiklabor hatte im Jahr 2003 den trinationalen Bildungspreis Medida gewonnen und eignete sich daher aus Sicht der Lehrenden Prof. Dr. Rendtel und Dr. Amit Ghosh gut, um sowohl in der Lehre als auch in der Prüfungssituation eingesetzt zu werden (vgl. Ghosh & Rendtel, 2008, S. 163). Darüber hinaus konnten die Prüfungsergebnisse schneller bewertet werden.

Als niedrigschwelliges Einstiegsszenario in computergestützte Prüfungen wurde damals eine Open-Book-Klausur durchgeführt. Open-Book bedeutet, dass Studierende alle Unterlagen und Aufzeichnungen aus der Lehre auch in der Prüfungssituation verwenden können (vgl. Schulz & Apostolopoulos, 2014). In solchen Prüfungsszenarien geht es weniger darum, Faktenwissen zu prüfen als darum, das bereits erworbene Faktenwissen als Syntheseleistung auf einen Fall anzuwenden.

Prüfungen in PC-Pools und der Mensa

Das Szenario der Open-Book-Prüfung hatte auch den Vorteil, dass studentische Notebooks eingesetzt werden konnten, auf die die Freie Universität aus Gründen der damaligen technischen Unterversorgung in den PC-Pools angewiesen war. Aufgrund der Konzeption als Open-Book-Prüfung mussten die Notebooks nicht zeitaufwändig vor der Prüfung auf unerlaubte Materialien untersucht werden, denn es war den Studierenden erlaubt, alle Materialien zu verwenden. Nur miteinander kommunizieren durften sie auch während einer solchen Prüfung nicht. Dass die Studierenden sich dennoch auf die Prüfung vorbereiten mussten, ist selbstredend, denn sonst hätten sie ihre Materialien nicht auf die zu bearbeitenden Fälle anwenden können.

Da es zu dieser Zeit noch keine separaten Räume für Notebook-Prüfungen gab, mussten die Studierenden mit eigenen Notebooks die Prüfung in der gesperrten Mensa des Fachbereichs Wirtschaftswissenschaft absolvieren. Die Abgabe der Ergebnisse am Ende der Prüfungen erfolgte mithilfe von USB-Sticks. Die beiden Lehrenden und weitere sechs Helfer sammelten umständlich von jedem studentischen Notebook die Ergebnisse einzeln ein.

Lessons Learnt

Aus den ersten durchgeführten paradigmatischen Prüfungen im Jahre 2005 konnten folgende Erkenntnisse extrahiert werden:

- Open-Book-Prüfungen eignen sich gut für einen *niedrigschwelligen Einstieg* in das Thema computergestützter Prüfungen.
- Unter der Prämisse der Open-Book-Prüfungen ist der Einsatz von studentischen Notebooks mit Kohorten von bis zu 150 Teilnehmenden *technisch und logistisch gut handhabbar*, da die einzuhaltenden technischen Sicherheitsbedingungen geringer als bei Closed-Book-Prüfungen sind.
- Die Abgabe bzw. das Einsammeln der Prüfungsergebnisse mittels USB-Sticks ist machbar, aber bei Kohortengrößen von 150 Teilnehmenden oder mehr zeitlich sehr *aufwändig*.

Prüfungsräume

Die Durchführung der Prüfungen in den PC-Pools des Rechenzentrums und den Mensa verlief problemlos. Es zeigte sich jedoch bereits Mitte 2006, dass dieses Verfahren nur bedingt skalierbar auf weitere Fachbereiche der Hochschule sein würde. Die technische Infrastruktur der Mensa war eben nicht auf computergestützte Prüfungsbetriebe vorbereitet worden und weitere Schritte zur Automatisierung des Prüfungsablaufs konnten somit nicht sofort realisiert werden. In der Folge wurden Überlegungen angestellt, ob die Räume der Mensa mit benötigter Strom- und Netzwerkinfrastruktur nachgerüstet werden könnten. Jedoch ergaben Schätzungen, dass sowohl die Kosten für die Nachrüstung der technischen Infrastruktur z. B. aufgrund fehlender Unterkellerung unter der Mensa als auch das Risiko, dass die dann installierte technische Infrastruktur vom normalen Mensabetrieb in Mitleidenschaft gezogen werden könnte (z. B. Essensreste in den Netzwerksteckdosen etc.), zu hoch sein würden. Stattdessen wurde von CeDiS und dem Rechenzentrum dafür votiert, normale Seminarräume, die in räumlicher Nähe zu den weiterhin für Prüfungen verwendeten zentralen PC-Pools standen, mit entsprechender technischer Infrastruktur für Prüfungen mit Notebooks auszustatten.

Prüfungssoftware

Die erfolgreichen Pilotversuche mit computergestützten Prüfungen sprachen sich in der Hochschule unter den Lehrenden schnell herum und ließen die Nachfrage bei CeDiS ansteigen. Daher musste schnellstmöglich eine zeitlich effizientere und für andere Fachbereiche generalisierbare Lösung für das Abgeben der Prüfungsergebnisse als das bislang praktizierte Einsammeln mit USB-Sticks gefunden werden. Nicht nur das Statistikkolabor sollte verwendet werden können, sondern auch weitere Standard-Applikationen wie Microsoft Word oder Excel. CeDiS entwickelte daher ab

Ende 2006 ein allgemeines Abgabetool, das die von den Studierenden bearbeiteten Prüfungsteile nach der Prüfung semi-automatisiert via WebDAV in einen gesicherten Bereich des Learning-Management-Systems Blackboard (LMS) transferieren konnte.

The screenshot shows a web browser window titled "CeDiS Klausurabgabe 1.1.3". The interface is divided into several sections:

- Nutzerdaten:** Contains three input fields: "Matrikelnr.", "Benutzer", and "Kennwort". There is a "Neu eingeben" button next to the "Matrikelnr." field and an "Anmelden" button next to the "Kennwort" field. The CeDiS logo is visible on the right.
- Aufgaben abgeben:** Contains three rows, each with a dropdown menu, a "Wähle Aufgabe" button, and a checked checkbox followed by "Aufgabe X abgeben".
- Abgabe der Aufgaben:** Contains two checkboxes: "Ja, ich habe alle Anwendungen geschlossen" and "Ja, ich möchte die oben angegebenen Aufgaben abgeben". There is an "Aufgaben abgeben" button to the right.
- Status:** A text box containing the instruction "1. Bitte geben Sie Ihre Nutzerdaten ein:" followed by three lines: "--> Matrikelnummer", "--> Benutzername", and "--> Kennwort".
- Programm Beenden:** A button at the bottom left.

Abbildung 1: Das CeDiS-Klausurabgabetool

Das Projekt „FU E-Examinations“ 2007–2010

Prüfungen in Notebook-Prüfungsräumen und PC-Pools



Abbildung 2: Statistikprüfung im Notebook-Prüfungsraum J27/14 der Freien Universität im Jahre 2008

Aufgrund der positiven Erfahrungen der computergestützten Prüfungen im Bereich der wirtschaftswissenschaftlichen Statistik beantragte CeDiS bei dem inneruniversitären Lenkungsgremium E-Learning (LGeL) das Projekt „FU E-Examinations – Digitaler Lehr- und Prüfungssaal.“ Das Projekt war zunächst befristet auf die Zeit von Anfang 2007 bis Ende 2008 und wurde dann bis Ende 2010 verlängert.

Ziel der ersten Stufe des Projekts von 2007 bis Ende 2008 war es, die in den Vorjahren gewonnen didaktischen, logistischen und technischen Erfahrungen mit computergestützten Prüfungen im Bereich der Statistik zu verallgemeinern und weiteren Fachbereichen zugänglich zu machen. Zu diesem Zeitpunkt gingen wir bei CeDiS noch davon aus, dass studentische Notebooks für computergestützte Prüfungen die erste Wahl wären. Daher wurden innerhalb des Projekts zunächst vier Seminarräume in der Silberlaube der Freien Universität zu Prüfungsräumen für studentische Notebooks aufgerüstet, indem in die Tische Netzwerk- und Stromanschlüsse eingebaut wurden.

Neben der Einrichtung von Notebook-Prüfungsräumen wurden universitätsweit einsetzbare Prüfungssoftware-Lösungen evaluiert, da es absehbar war, dass die langfristige Weiterentwicklung des Abgabetools zu kosten- und zeitintensiv und perspektivisch nicht ausreichend flexibel sein würde, um die Bedarfe der unterschiedlichen Fachbereiche zu decken. Es sollten schließlich nicht nur Open-Book-

Prüfungen computergestützt möglich sein, sondern auch Closed-Book-Prüfungen, bei denen Studierende Faktenwissen reproduzieren und demzufolge keinerlei eigene Unterlagen oder Aufzeichnungen verwendet werden sollten. Closed-Book-Prüfungen waren jedoch auf studentischen Notebooks alleine mit dem Abgabetool noch nicht umsetzbar.

Ziel der zweiten Stufe des Projekts von 2009 bis Ende 2010 war es dann, aus den evaluierten Prüfungsplattformen eine für den universitätsweiten Einsatz auszuwählen und E-Examinations aus dem Projektstatus in einen zentralen Service für Lehrende der Universität zu überführen.

Notebook-Prüfungsräume

Die Kapazität der Notebook-Prüfungsräume betrug nach Fertigstellung im Sommersemester 2007 insgesamt 221 Plätze. Zusammen mit den PC-Pools des Rechenzentrums (196 Plätze verteilt auf 7 Räume) standen somit rechnerisch ab dem Sommersemester 2007 in insgesamt 12 Räumen bis zu 417 Plätze für computergestützte Prüfungen zur Verfügung. Aus Gründen einer geteilten Nutzungskonzeption konnten diese Kapazitäten jedoch nie voll ausgeschöpft werden. Die PC-Pools des Rechenzentrums wurden parallel auch für universitäre Schulungen verwendet und so ergaben sich zu häufige logistische Engpässe qua zeitlicher Überschneidungen. Erschwerend kam hinzu, dass universitätsweit nur ca. 53% der Studierenden über mit dem Universitätsnetzwerk kompatible Notebooks verfügten (vgl. Reinhardt, 2006).

Um die Notebook-Verbreitung bei den Studierenden der Freien Universität zu erhöhen und gleichzeitig die Heterogenität der von den Studierenden verwendeten Notebooks zu senken, brachte CeDiS gemeinsam mit dem universitären Rechenzentrum Kooperationen mit Hardware-Herstellern auf den Weg. Die damals ausgewählten Hardware-Hersteller Dell und Apple räumten über nur aus dem Universitätsnetzwerk zugänglichen Webshops Vergünstigungen zwischen 15 und 20% auf ihre Geräte ein.

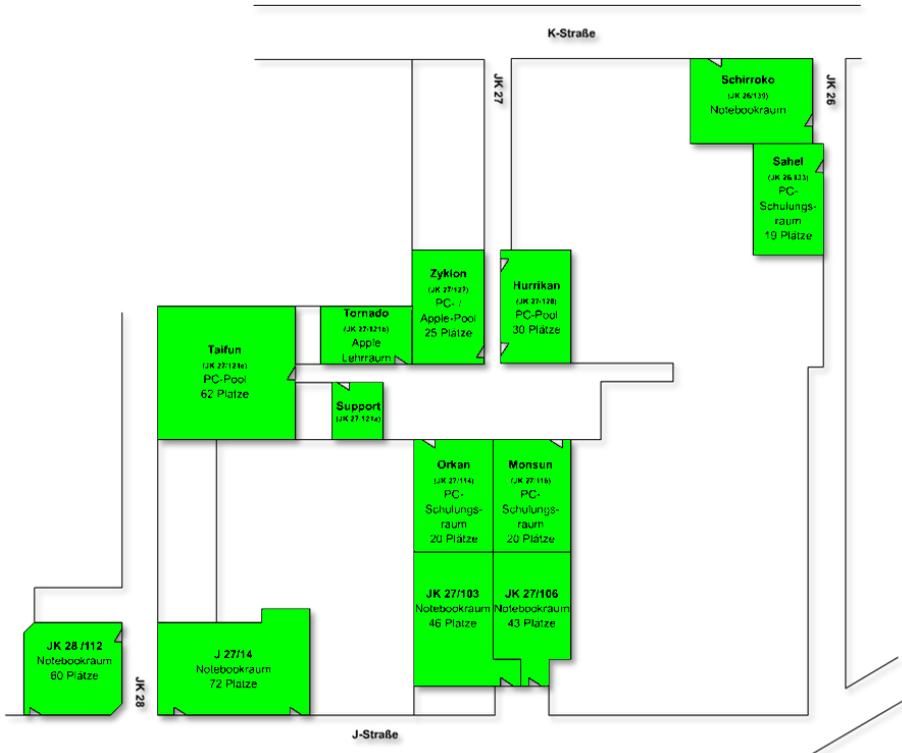


Abbildung 3: Notebook-Prüfungsräume und zentrale PC-Pools des Rechenzentrums der Freien Universität im Jahre 2008

Um die Kompatibilität der studentischen Notebooks massenhaft schneller und einfacher als durch universitäre Mitarbeiter testen zu können, hatte CeDiS bereits 2007 begonnen, ein Kompatibilitätstool zu entwickeln, das die Studierenden von der CeDiS E-Examinations Website herunterladen und auf ihre Notebooks installieren konnten. Das Kompatibilitätstool überprüfte, ob die Netzwerkeinstellungen akkurat und die erforderlichen Applikationen installiert waren.

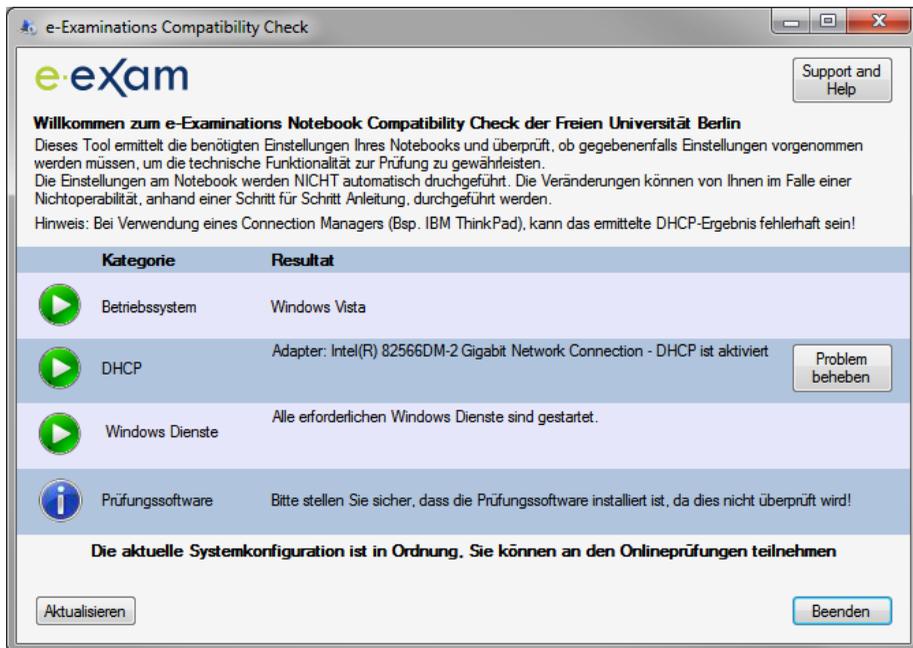


Abbildung 4: Das Kompatibilitätstool im Jahre 2009

Prüfungssoftware

Der Markt für computergestützte Prüfungslösungen war in den Jahren 2007 und 2008 recht überschaubar. Neben Questionmark hatten nur LPLUS und ExamSoft ernstzunehmende universelle Prüfungsplattformen im Angebot. Die Prüfungsfunktionalitäten in den Learning-Management-Systemen (LMS) waren demgegenüber noch weitestgehend rudimentär. Sowohl das kommerzielle LMS Blackboard als auch die Open-Source-LMS wie Moodle oder ILIAS beinhalteten weitestgehend einfache Tests und Quizzes.

Der Unterschied zwischen Blackboard und den Open-Source-LMSen bestand aus Sicht der Freien Universität lediglich darin, dass Blackboard als LMS bereits als universitäre Installation im produktiven Einsatz war und damit kostengünstig auch für Prüfungen hätte eingesetzt werden können. Da das von CeDiS entwickelte Abgabewerkzeug darüber hinaus auch Blackboard verwendete, wurde Blackboard weiterhin als mögliche universitätsweite Prüfungsplattform in Erwägung gezogen, die mittels zu kaufender und selbst zu entwickelnder Erweiterungen die Bedarfe der Freien Universität hätte decken können.

Der Hauptfokus der Evaluation der reinen Prüfungsplattformen lag auf den kommerziellen Prüfungsplattformen Perception (Questionmark), dem TestStudio (LPLUS) und ExamIntelligence (ExamSoft). Alle drei Produkte waren ähnlich konzipiert als Server-Client-Lösung. Sie waren webbrowsersbasiert, stellten umfangreiche Fragetypen zur Verfügung und wiesen eine ähnliche Usability bei akzeptabler Konfigurierbarkeit auf.

Folgende Anwendungen wurden für den Prüfungseinsatz evaluiert:

Tabelle 1: Übersicht evaluierter Softwarelösungen für den Prüfungseinsatz

Software	Typ	Installation	Source-Code einsehbar
Perception (Questionmark)	All-in-One	Universitäre Installation möglich	Nein
TestStudio (LPLUS)	All-in-One	Universitäre Installation möglich	Nur unter NDA ¹
Exam Intelligence (ExamSoft)	All-in-One	Nur Online (Cloud)	Nein
Lockdown Browser (Respondus)	All-in-One	Universitäre Installation vorhanden (LMS Blackboard)	Nein
Exam 4 (Extegrity)	Juristische Examina	Nur Online (Cloud)	Nein

Parallel zu den Evaluationen der Prüfungslösungen wurde ab dem Sommersemester 2007 für die Wirtschaftsstatistik das von CeDiS entwickelte Abgabetool zum Einsammeln der absolvierten Prüfungsaufgaben eingesetzt. Damit Studierende, die mit eigenen Notebooks an der Prüfung teilnahmen, das Abgabetool verwenden konnten, mussten sie es im Vorfeld auf ihren Notebooks installieren. Während der Prüfung erhielten sie einen einmalig gültigen User-Namen und ein Passwort. In das Abgabetool wurden diverse Sicherheitsfunktionen integriert, z. B. um zweifache Abgaben oder doppelte Logins zu verhindern. Ferner deaktivierte das Abgabetool direkt nach der vollendeten Abgabe das Zugangskonto, sodass ein wiederholtes Einloggen des gleichen Studierenden nicht möglich war. Nach der Prüfung konnten die Lehrenden dann aus dem LMS die computergestützt bearbeiteten Prüfungsteile aller Studierenden herunterladen und diese auf ihren lokalen PCs bewerten.

¹ Ein sog. Non-Disclosure-Agreement (NDA) ist eine vertragliche Vereinbarung zwischen Partnern, über die ausgehandelten Details – meist auch über ein mögliches Vertragsende hinaus – grundsätzlich Stillschweigen zu bewahren, weil hier meist Geschäftsgeheimnisse betroffen sind, die für Mitbewerber von großem Interesse sein könnten. In Fall von LPLUS erlaubt die geschlossene NDA-Vereinbarung der Freien Universität Einsicht in den Source-Code der Prüfungsplattform LPLUS TestStudio.

Lessons Learnt

Am Ende der Projektphase im Jahre 2010 konnten folgende Punkte festgehalten werden:

- Notebook-Prüfungsräume bedeuteten faktisch *geringe Hardwarekosten* für die Universität.
- Insgesamt 12 Räume, in denen computergestützte Großprüfungen durchgeführt wurden, erforderten allerdings *zu viel didaktisches und zu viel technisches Aufsichtspersonal*.
- Die Heterogenität der Studierenden-Notebooks bedingte einen *hohen technischen Support-Aufwand*.
- Trotz der Kooperation von Hochschule und Notebook-Herstellern griffen Studierende eher zu *kostengünstigeren Notebook-Angeboten der Discounter*.
- Computergestützte Prüfungen wurden auch von Lehrenden nachgefragt, die zuvor noch keine E-Learning-Instrumente in der Lehre einsetzten.
- Es gab nur *wenige ausgereifte Prüfungssoftware-Lösungen*.
- Computergestützte Prüfungen *verkürzen die Bewertungsphase* von Prüfungen.

Notebook-Prüfungsräume

Ogleich der Ansatz, studentische Notebooks in die Prüfung einzubeziehen, zunächst geringe Investitionskosten versprach, da von Seiten der Hochschule nur wenig Geld in die Infrastruktur investiert werden musste, um Seminarräume notebookprüfungskompatibel herzurichten, zeigte sich, dass der Support-Aufwand, um die heterogene Masse studentischer Notebooks zu konfigurieren, insgesamt hoch war und damit die verdeckten Kosten für Personal höher als erwartet waren. Zwar war von Vorteil, dass die Lehrenden der Wirtschaftsstatistik sehr technikaffin waren und viele kleinere Konfigurationsfragen der Studierenden schon in den Seminaren lösen konnten, wodurch nur wenige Studierende vom Fachbereich an CeDiS als Supportfälle weitergeleitet werden mussten. Ein generalisierbares und für die gesamte Hochschule anwendbares Verfahren stellte dies jedoch nicht dar. Nur wenige Lehrende aus anderen Fachbereichen hatten eine ähnliche technische Expertise. Darüber hinaus erwiesen sich trotz der Vorkehrungen einige Notebooks von Studierenden erst in einem der Notebook-Prüfungsräume als inkompatibel. Für diese wurden schon in der ersten Stufe des Projekts ab 2007 zwei mögliche Fallback-Szenarien vorbereitet. Entweder konnten sie sich entscheiden, die Prüfung im PC-Pool durchzuführen oder sie absolvierten die Prüfung dennoch am eigenen Notebook und die Ergebnisse wurden vom Notebook nach der Prüfung ersatzweise mittels USB-Stick eingesammelt. Insgesamt wurde das Prozedere der Nutzung

studentischer Notebooks von CeDiS zwar als machbar, aber mittelfristig als nicht ökonomisch eingestuft.



Abbildung 5: Notebook-Prüfungsraum J 27/14 der Freien Universität im Jahre 2008

Die Kooperation mit den Hardwareherstellern, die einerseits die Heterogenität der unter den Studierenden verbreiteten Notebooks senken sollte, andererseits aber auch die universitätsweite Verbreitung von Notebooks unter den Studierenden heben sollte, wirkte nur marginal. Bereits Ende 2009 zeigte sich, dass nur wenige Studierende die im Rahmen der Kooperation mit Dell und Apple vergünstigten Notebook-Modelle erworben hatten. In den Notebook-Prüfungsräumen nahm die Heterogenität der studentischen Notebooks nicht ab.

In Gesprächen mit Studierenden der Wirtschaftswissenschaft zeigte sich, dass diese die Angebote aus den Kooperationen zwar befürworteten, sie den Preis (zwischen 800 und 1200 €) für die Modelle von Dell und Apple im Vergleich zu Notebook-Angeboten von den Discontnern (zwischen 500 und 700 €) aber als immer noch zu hoch für das knappe studentische Budget betrachteten. Eine Senkung der Heterogenität der studentischen Notebooks konnte auf diesem Weg nicht erreicht werden.

In der zweiten Phase ab 2009 stand dann das von CeDiS entwickelte Kompatibilitätstool zum Testen der Kompatibilität der studentischen Notebooks mit dem Universitätsnetzwerk zur Verfügung. Es funktionierte weitestgehend erwartungsgemäß. Studierende luden das Tool herunter und folgten den angezeigten Anweisungen. Sofern das Tool meldete, dass das Notebook nicht kompatibel ist, hatten die Studierenden zwei Möglichkeiten: entweder konnten sie ihr Notebook prüfungskompatibel umkonfigurieren – was selten gemacht wurde, da dann häufig Schwierigkeiten mit dem privaten DSL-Anschluss entstanden – oder sich einen Prüfungsplatz in den PC-Pools reservieren lassen und die Prüfung ohne eigenes Notebook absolvieren. Etwa 7% der vom Kompatibilitätstool als funktionsfähig eingestuften studentischen Notebooks funktionierten dennoch nicht in den Notebook-Prüfungsräumen. Den betreffenden Studierenden stand in diesen Fällen weiterhin die oben bereits beschriebene Möglichkeit offen, die Prüfung statt am eigenen Notebook in den universitären PC-Pools zu absolvieren oder sie am eigenen Notebook abzulegen und sie hinterher mittels USB-Stick einsammeln zu lassen.

Der Einsatz des Kompatibilitätstools verbesserte zwar die Prognostizierbarkeit der Anzahl der inkompatiblen Notebooks, der Support-Aufwand und sowohl die logistische (Umsetzen der Teilnehmenden in PC-Pools) als auch technische Fallbacklösungen (Einsammeln der Ergebnisse mittels USB-Sticks) wurden jedoch beibehalten, auch wenn die Anzahl der Fälle sank.

In der Gesamtbetrachtung Ende 2009 zeigte sich, dass das Kompatibilitätstool äußerst wartungsaufwändig war. Grund war wieder einmal die Heterogenität der Hardware der unter den Studierenden verbreiteten Notebooks. Die Vielzahl unterschiedlicher Windows Versionen (Windows 2000, Windows XP 32/64bit, Windows Vista 32/64bit, Windows 7 32/64bit) mit unterschiedlichen Ständen der installierten Service-Packs und die mannigfaltigen Chipsätze der eingebauten Netzwerkkarten erforderten häufige Aktualisierungen des Kompatibilitätstools. Ende 2010 wurde es letztmalig aktualisiert und die Entwicklung eingestellt.

Prüfungssoftware

Im direkten Vergleich zwischen LMSen und genuinen Prüfungsplattformen zeigte sich, dass der Funktionsumfang und die Sicherheitsüberlegungen bei Prüfungsplattformen einen erheblich höheren Stellenwert hatten als LMSen.

Für Blackboard gab es zu diesem Zeitpunkt nur wenige Erweiterungen auf dem Markt, die für Prüfungen geeignet waren. Dazu gehörte z. B. der Respondus Lock-down Browser. Der Aufwand, diese zu zukaufenden Erweiterungen für einen universitätsweiten Einsatz zu konfigurieren und ggf. weitere Erweiterungen in Eigenregie unter der Ägide von CeDiS zu entwickeln oder entwickeln zu lassen, wurde unter Aspekten der Nachhaltigkeit und Wartbarkeit als mittelfristig ungewisser Kostenfaktor eingeschätzt und als Ansatz deshalb wieder verworfen.

Ferner waren in Prüfungsplattformen technische Sicherheitsüberlegungen wie redundante Speicherungen von Prüfungsdaten grundlegend in das Programmdesign einbezogen worden. Während LMSe datenbankseitig auf viele parallele Lesezugriffe optimiert waren, waren Prüfungsplattformen schon damals auf viele parallele Schreibzugriffe optimiert.

Das Vorhalten hochsensibler personenbezogener Prüfungsdaten auf Cloud-Servern mit unbekanntem weltweiten Standorten war von CeDiS und von Seiten der Datenschutzbeauftragten der Freien Universität ausgeschlossen worden, wodurch Prüfungsplattformen wie ExamIntelligence oder Exam 4 trotz der akzeptablen Funktionsumfänge nicht tiefenanalysiert wurden. Die Datenschutzbeauftragte empfahl stattdessen eine Installation auf Servern im Datacenter der Hochschule. Die Möglichkeit einer Installation der Prüfungsplattform im universitären Rechenzentrum wurde nur von Questionmark Perception und dem LPLUS TestStudio angeboten.

Der Funktionsumfang von Questionmark Perception und dem LPLUS-TestStudio war zu dieser Zeit bereits recht hoch und weitestgehend vergleichbar. Zu den Aufgabentypen gehörten u. a. Multiple-Choice, Single-Choice, Drag-and-Drop, Freitexteingaben und Lückentexte. Um auch Closed-Book-Prüfungen realisieren zu können, boten sowohl Questionmark als auch LPLUS Sicherheitsbrowser an, mit denen studentische Notebooks während der Prüfungssituation in einen gesicherten Prüfungsmodus versetzt werden konnten. Im gesicherten Prüfungsmodus waren sämtliche Daten auf dem Notebook während der Durchführung einer Prüfung nicht zugänglich und nur die Prüfung sichtbar. Den wesentlichen Unterschied der beiden Produkte machten folgende für die Freie Universität wichtige Punkte: Erstens bot LPLUS als Hersteller des TestStudios die Möglichkeit an, mittels eines Non-Disclosure-Agreements Einsicht in den Source-Code des TestStudios zu erhalten. Zweitens war das TestStudio die einzige Prüfungsplattform, die eine permanente Zwischenspeicherung von langen Freitextantworten (in Form von einzelnen separaten Text-Revisionen, die alle 30 Sekunden auf dem Server gespeichert werden) beinhaltete. Da die Freie Universität eine überwiegend geistes- und sozialwissenschaftlich ausgerichtete Hochschule ist und häufig längere Antworttexte von den Studierenden verfasst werden müssen, wurde dieses Sicherheitsfeature für Havarie-Situationen (z. B. Rechnerabsturz, Stromausfall, Netzwerkunterbrechungen etc.) als „Must-Have“ eingestuft.

Verkürzung der Bewertungsphasen

Die mit computergestützten Prüfungen einhergehende Arbeitshypothese, dass sie das Potenzial zur Zeitersparnis in der Bewertungsphase am besten bei Multiple-Choice-Prüfungen entfalten können, wurde im Sommersemester 2009 einer genaueren Betrachtung unterzogen (vgl. Schulz & Apostolopoulos, 2011, S. 38 ff.).

Komparative Messungen der Bewertungsphase zeigten, dass zwar die *prozentuale* Zeitersparnis bei Multi-Choice-Prüfungen erheblich höher war als bei Freitextprüfungen, bei denen Studierende Texte verfassen mussten, die *effektive* Zeitersparnis unter Einbezug des Faktors der Personalkosten jedoch höher ist, wenn schriftliche Freitextprüfungen zu computergestützten migriert werden.

Tabelle 2: *Effektive Zeitersparnis vs. prozentuale Zeitersparnis*

	Multiple-Choice-Prüfung	Freitextprüfung
Bewertungsdauer in Tagen ² pro 100 Prüfungsteilnehmende	computergestützt: 0,06 Tage schriftlich: 0,9 Tage	computergestützt: 11,5 Tage schriftlich: 17 Tage
Zeitersparnis in Prozent	93%	33%
Kostenersparnis ³ auf Basis der effektiven Zeit	215 €	1.400 €

Der Grund dafür liegt in der ohnehin schon hohen Zeiteffizienz von Multi-Choice-Prüfungen. Denn bei einer schriftlichen Multi-Choice-Prüfung ist die Bewertungsphase qua hoher Standardisierung ohnehin sehr kurz. Anders verhält es sich bei Freitextprüfungen. Die Bewertungsphase dieser Prüfungen ist erheblich zeitintensiver, da die Texte von den Lehrenden gelesen und dann bewertet werden. Vermutlich entfalten computergestützte Freitextprüfungen eine höhere *effektive* Zeitersparnis wegen der besseren Lesbarkeit der am Computer verfassten Texte gegenüber handschriftlichen Texten.

² Bei einer rechnerischen täglichen Arbeitszeit von 7,3 h

³ Bei auf TV-L gemittelten täglichen Personalkosten von 250 €

Phase III (seit 2011)

Das Prüfungsszenario im Prüfungszentrum



Abbildung 6: Publizistik-Prüfung im E-Examination Center der Freien Universität im Jahre 2013

Nachdem CeDiS in den vorvergangenen Jahren mehrere Szenarien für computergestützte Prüfungen mit kleinen und großen Kohorten mit mehreren hundert Studierenden sowohl mit studentischen Notebooks als auch mit von der Hochschule administrierten PC-Pools erprobt hatte, entschloss sich die Universitätsleitung, den Arbeitsbereich E-Examinations ab 2011 auszuweiten und die E-Examination-Services als einen universitätsweit angebotenen Dienst zu institutionalisieren.

Die Strategie, computergestützte Prüfungen mittels studentischer Notebooks massenhaft zum Einsatz zu bringen, stieß seit Ende 2009 an Grenzen. Obwohl die Nachfrage nach computergestützten Prüfungen von Seiten der Lehrenden grundsätzlich anstieg, nahm gleichzeitig die Verbreitung im Jahre 2010 nur wenig zu. CeDiS konnte dafür folgende Gründe identifizieren: Studentische Notebooks konnten zwar prüfungskompatibel konfiguriert werden, aber ein erhöhter personeller Support-Aufwand durch CeDiS-Mitarbeiter/innen war dennoch sowohl im Vorfeld von Prüfungen als auch während der Prüfungsdurchführung erforderlich. Der erhöhte technische Aufwand, der aus der Heterogenität der studentischen Notebooks resultierte, führte wiederum dazu, dass Lehrende dem Konzept, studentische Notebooks zu verwenden, ambivalent gegenüberstanden. Von an der Freien Universität lehrenden Juristen wurde darüber hinaus die Frage formuliert, ob mit technisch unterschiedlich ausgestatteten studentischen Notebooks das Gebot der Gleichbehandlung aller Studierenden in einer Prüfungssituation überhaupt erfüllt werden könnte.

Diese Erfahrungen und Einschätzungen führten dazu, dass sich die Universitätsleitung für die Einrichtung eines vollausgestatteten Raums für computergestützte Massenprüfungen entschloss. Der Arbeitsauftrag an CeDiS lautete 2011:

- ein technisch vollausgestattetes *Prüfungszentrum* für Großkohorten zu konzipieren und anschließend dauerhaft zu betreiben, um sowohl *die technischen Rüstzeiten und den Support-Aufwand* als auch den *personellen Aufwand für fachdidaktisches Aufsichtspersonal* gering zu halten,
- eine universitätsweit einsetzbare *Prüfungssoftware-Lösung* anzuschaffen und dauerhaft zu betreiben,
- in Zusammenarbeit mit der Rechtsabteilung der Universität *rechtlich belastbare Rahmenbedingungen* für computergestützte Prüfungen in der Universität zu schaffen und fortlaufend zu aktualisieren und
- die *inneruniversitäre Nachfrage* nach E-Examinations sukzessive zu decken.

Prüfungsraum: Das E-Examination Center (EEC)

Ab Ende 2011 wurde begonnen, das E-Examination Center (EEC) in einem ehemaligen Großraumlabor des Instituts für anorganische Chemie einzurichten. Die Inbetriebnahme erfolgte im Februar 2013 mit einer Prüfung bei dem damaligen Vizepräsidenten der Freien Universität Prof. Dr. Klaus Beck. Das seitdem im Einsatz befindliche EEC bietet Platz für bis zu 151 Studierende und wird überwiegend für computergestützte Prüfungen und in seltenen Fällen für inneruniversitäre Schulungen mit hohen Teilnehmerzahlen verwendet. Die PCs im EEC sind für Prüfungen vorkonfiguriert. Neben den im universitären Rechenzentrum gehosteten Prüfungsservern sind keine weiteren Internetdienste im EEC erreichbar. Der Raum ist akustisch optimiert, sodass Prüfungsteilnehmer sich bestmöglich konzentrieren können. Zwischen Februar 2013 und Juli 2015 wurden im EEC insgesamt mehr als 25.000 Studierende bzw. Studienplatzbewerber computergestützt geprüft.

Neben den inneruniversitären Prüfungen kooperiert der Arbeitsbereich E-Examinations im Bereich der Weiterentwicklung der Nutzungsszenarien des E-Examination Centers in Richtung (eignungs-)diagnostischer Tests und Prüfungen mit der Charité Universitätsmedizin Berlin und der Senatsverwaltung für Inneres und Sport.

n = 5.837 Prüfungsteilnehmende

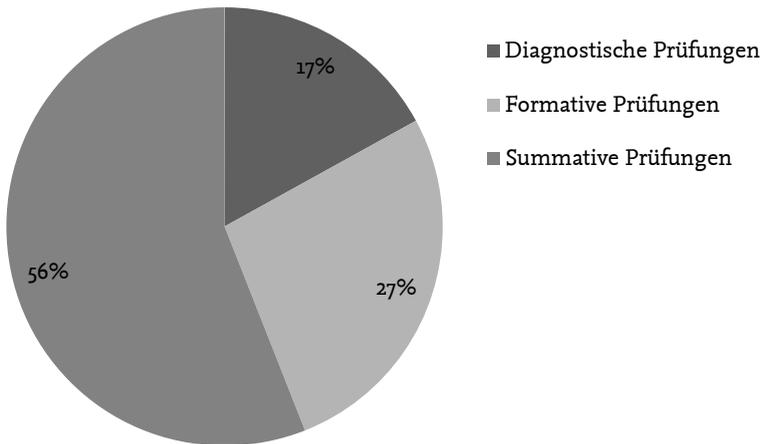


Abbildung 7: Verteilung der im EEC durchgeführten Prüfungstypen im Zeitraum Januar bis Juni 2015

Nach der Eröffnung des E-Examination Centers (EEC) und der Anschaffung von Prüfungsservern lässt sich folgendes Fazit ziehen. Das EEC senkt die Vorbereitungs- und Rüstzeiten gegenüber Notebook-Prüfungsräumen deutlich. Der personelle Aufwand für die Anlagensteuerung und für die Wartung der eingesetzten PCs ist gering. Unterschiedliche Prüfungsmodi können on-the-fly umgeschaltet werden. Sofern erforderlich kann in Prüfungen Drittsoftware (z. B. Microsoft Access, Microsoft Excel, R etc.) verwendet werden. Die Rüstzeiten (logistische Ein- und Auslasszeit von bis zu 151 Studierenden, kurze Einführung in die Bedienung der Prüfungsplattform und Sicherheitshinweise für Havarie-situationen) zwischen zwei Prüfungen betragen in der Regel 45 Minuten.

Der initiale Investitionsaufwand, um ein computergestütztes Prüfungszentrum zu eröffnen, ist jedoch sehr hoch und es sind Folgeinvestitionen in die PCs erforderlich, wenn diese veraltet sein werden.

Für das EEC wurde vom Arbeitsbereich E-Examinations im Vorfeld ein Betriebskonzept entwickelt. Derzeit sind zwei Angestellte in Vollzeit und zwei Studierende beschäftigt. Eine Erweiterung auf vier studentische Hilfskräfte ist bis Anfang 2016

vorgesehen. Im Betriebskonzept ist ausgeführt, wie die unterschiedlichen Prüfungstypen - diagnostische, formative und summative Prüfungen (vgl. Crisp, 2009) - durchgeführt werden können.

Prüfungssoftware

Neben der Einrichtung des Prüfungsraums wurde ab 2011 in enger Zusammenarbeit mit dem Hersteller der Prüfungssoftware LPLUS ein Hochverfügbarkeitscluster konzipiert. Der Cluster wurde auf die Durchführung von bis zu 800 parallelen computergestützten Prüfungen ausgelegt. Es wurde neben dem physischen Produktivsystem (zwei Application-Server und drei Datenbank-Server) ein äquivalent konzipiertes Testsystem (vollständig virtualisiert) eingerichtet. Das virtualisierte Testsystem dient dem Testen von Software-Updates und Patches, bevor sie auf dem Produktivsystem installiert werden. Der Prüfungscluster wurde Mitte 2012 in den Produktiveinsatz genommen.

Rechtliche Grundlagen

Auf Grundlage der Einschätzung der Verwaltungsrichter Niehues und Fischer, dass „die *elektronische* Prüfung [...] in diesem Zusammenhang nicht als eine Art schriftliche Prüfung zu verstehen [ist]“ (Niehus & Fischer, 2010, S.12, Hervorh. im Original), wurde von CeDiS und der Rechtsabteilung die Rahmenstudienprüfungsordnung (RSPO) um computergestützte Prüfungen ergänzt. Die RSPO trat im März 2013 in Kraft. Fachbereiche können seitdem mithilfe standardisierter, von der Rechtsabteilung erarbeiteter Formulierungen in ihren Prüfungsordnungen auf die RSPO verweisen, um computergestützte Prüfungen einzusetzen.

Evaluationsergebnisse

Im Zeitraum 01.02.2014 bis 31.05.2015 wurden von der inneruniversitären Arbeitsstelle Lehr- und Studienqualität in Zusammenarbeit mit CeDiS begleitende Evaluationen zur Computerisierung des Prüfungswesens im EEC durchgeführt. Im insgesamt 16 Monate umfassenden Evaluationszeitraum wurden 15.465 Teilnehmende im EEC computergestützt geprüft. Die Evaluation fand nach der Prüfung am gleichen Computer statt, an dem auch die Prüfung absolviert wurde. Die Teilnahme an der Evaluation war für Studierende *nicht* obligatorisch. Es nahmen insgesamt 1.309 Studierende (8,5%) an den Evaluationen teil. Das Durchschnittsalter der befragten Studierenden betrug 25 Jahre. Rund 64% der Befragten waren weibliche und 36% männliche Studierende. Die Befragten studierten Fächer aus den Bereichen Naturwissenschaften (inkl. Human- und Veterinärmedizin), Sozial- und Geisteswissenschaften. Im Folgenden finden sich Auszüge der noch separat zu publizierenden vollständigen Evaluationsergebnisse.

Anzahl absolvierter computergestützter Prüfungen

Für eine Mehrheit von 43% der befragten Studierenden war dies die erste absolvierte computergestützte Prüfung. Für 23% der befragten Studierenden war dies bereits die zweite computergestützte Prüfung und 32% der Teilnehmenden hatte bereits zwischen drei und sechs computergestützte Prüfungen absolviert. Ein bislang nur kleiner Teil der Befragten von 2% hatte bereits mehr als sechs computergestützte Prüfungen abgelegt.

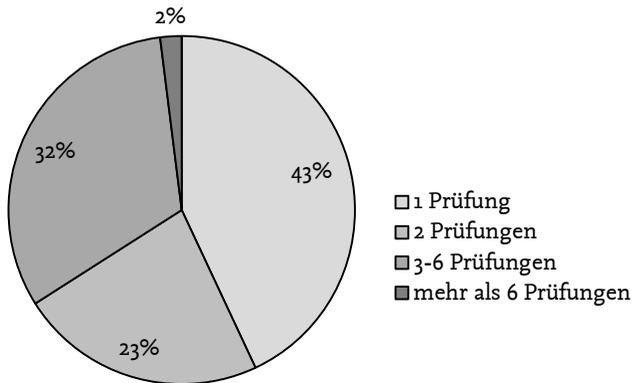


Abbildung 8: Prozentuale Verteilung der Anzahl absolvierter computergestützter Prüfungen

Zufriedenheit mit den Rahmenbedingungen im E-Examination Center

Das E-Examination Center (EEC) ist in der Konzeptionsphase auf Ausbreitung des Schalls und auf günstige Klimatisierungsverhältnisse optimiert worden, sodass Studierende sich in diesem Raum bestmöglich konzentrieren konnten. In der Evaluation zeigte sich, dass die Studierenden mit diesen räumlichen Bedingungen in hohem Maße zufrieden waren.

Bedienerfreundlichkeit der Prüfungssoftware

Insgesamt 98% der befragten Studierenden gaben an, dass sie mit der Bedienung des von der Freien Universität eingesetzten E-Examination-Systems (EES) sehr gut bzw. eher gut zurechtkamen. Nur 2% gaben an, dass die Software eher schlecht bzw. sehr schlecht zu bedienen wäre.

Vergleich klassisch schriftliche Prüfungen mit computergestützten Prüfungen

Obwohl es sich um ein recht neues Prüfungskonzept handelte, fanden nur 33% der befragten Studierenden computergestützte Prüfungen schlechter oder eher schlechter als klassisch schriftliche Prüfungen. Demgegenüber fanden insgesamt 67% computergestützte Prüfungen entweder besser als oder zumindest gleich gut wie klassisch schriftliche Prüfungen.

Korrelationen

Es zeigte sich, dass die subjektive Einschätzung des vermuteten Prüfungsergebnisses einen negativen Einfluss auf die Bewertung computergestützter Prüfungen hatte. D.h., wenn befragte Studierende davon ausgingen, dass sie ein schlechtes Prüfungsergebnis erreicht hatten, bewerteten sie computergestützte Prüfungen auch eher negativ.

Auch zeigten sich Gewöhnungseffekte. Befragte Studierende, die bereits eine höhere Anzahl an computergestützten Prüfungen absolviert hatten, gaben mit höherer Wahrscheinlichkeit an, diese gegenüber klassisch schriftlichen Prüfungen zu bevorzugen.

Fazit und Ausblick

Die Freie Universität hat einen mittlerweile zehnjährigen Weg im Bereich computergestützter Prüfungen hinter sich. Obgleich wir in der Frühphase computergestützte Prüfungen noch als eine Erweiterung von E-Learning-Szenarien eingestuft hatten, zeigte sich schon bald, dass sie qua anders gelagerter (prüfungs-)rechtlicher und technischer Anforderungen kein Thema sind, das nebenher behandelt werden kann und der Fokussierung bedarf.

Dennoch lohnen sich computergestützte Prüfungen. Sie erleichtern den Hochschulalltag von Lehrenden und Studierenden erheblich. Die Bewertungszeiten von Prüfungen werden deutlich verkürzt und Prüfungsergebnisse können erheblich früher veröffentlicht werden. Dies hilft der sowohl den Lehrenden, als auch der universitären Administration und den Studierenden. Auf didaktischer Ebene ergänzen computergestützte Prüfungen die moderne Hochschullehre um Technologien, die die Anschaulichkeit komplexer Inhalte und Themen erhöhen.

Die gegenwärtigen Entwicklungen zeigen, dass computergestützte Prüfungen flächendeckend an den deutschsprachigen Hochschulen thematisiert werden. Für die Freie Universität bedeutet dies, dass in den kommenden Jahren am Arbeitsbereich E-Examinations bei CeDiS rechtliche und didaktische Aspekte noch differenzierter bearbeitet werden müssen. Zu den rechtlichen Aspekten gehören die folgenden prüfungsrechtlichen und datenschutzrechtlichen Fragen:

- Wie können computergestützte Prüfungen in den Prüfungsordnungen der Hochschulen rechtlich belastbar verankert werden?
- Wie können computergestützte Prüfungen datenschutzrechtlich sicher und gleichzeitig nachhaltig digital aufbewahrt werden?

Im Bereich der Prüfungsdidaktik liegen im deutschsprachigen Raum bislang nur Arbeiten des Sprachwissenschaftlers Thomas Tinnefeld (2013) vor. Das Thema Prüfungsdidaktik sollte in den nächsten Jahren unter Einbezug der Messtheorie bzw. der Testtheorie für das Feld computergestützter Prüfungen stärker fruchtbar gemacht werden.

Referenzen

- Crisp, G. (2009). Interactive e-Assessment: Moving beyond multiple-choice questions. University of Adelaide, South Australia: Centre for Learning and Professional Development. Abgerufen unter http://www.academia.edu/1187193/Interactive_e-Assessment_moving_beyond_multiple-choice_questions [10.09.2015]
- Ghosh, A & Rendtel, U. (2008). Unterrichten und Prüfen mit dem Statistikkolabor: Ein Erfahrungsbericht. In R. Münnich (Hrsg.), *ASTA Wirtschafts- und Sozialstatistisches Archiv* (Bd. 2, S. 145-164). Berlin: Springer.
- Niehues, N. & Fischer, E. (2010). *Prüfungsrecht*. München: C. H. Beck.
- Palfrey, J. & Gasser, U. (2008). *Born Digital: Understanding the First Generation of Digital Natives*. New York: Basic Books.
- Schulz, A. & Apostolopoulos, N. (2010). FU E-Examinations: E-Prüfungen am eigenen Notebook an der Freien Universität Berlin. In C. Rüdell & S. Mandel (Hrsg.), *E-Assessment: Einsatzszenarien und Erfahrungen an Hochschulen*. (S. 23-46). Münster: Waxmann.
- Schulz, A. & Apostolopoulos, N. (2011). E-Examinations put to test: Potenziale computergestützter Prüfungen. *Hamburger eLearning Magazin*, 7, 37-39. Abgerufen unter <http://www.uni-hamburg.de/elearning/hamburger-elearning-magazin-07.pdf> [10.09.2015]
- Schulz, A. & Apostolopoulos, N. (2014). E-Examinations at a glance: Die Computerisierung des Prüfungswesens an der Freien Universität Berlin. In T. Skerlak, H. Kaufmann & G. Bachmann (Hrsg.), *Lernumgebungen an der Hochschule: Auf dem Weg zum Campus von Morgen*. (S. 283-298). Münster: Waxmann.

Reinhardt, J. (2006). Ergebnisse der Online-Umfrage zum E-Learning Einsatz an der Freien Universität und der Charité Universitätsmedizin Berlin. Abgerufen unter www.cedis.fu-berlin.de/e-learning/beratung/lehren_mit_neuen_medien/erfahrungen/umfrage_sommer06/ergebnisse/index.html [05.08.2015]

Tinnefeld, T. (2013). *Dimensionen der Prüfungsdidaktik: Analysen und Reflexionen zur Leistungsbewertung in den modernen Fremdsprachen*. Saarbrücken: Hochschule für Technik und Wirtschaft des Saarlandes.

Abgerufen unter <https://drive.google.com/file/d/0B-2HUXwM3-xiYXhXQndnSTVCCdG/edit?usp=sharing> [10.09.2015]

Vita

Alexander Schulz koordiniert den Arbeitsbereich E-Examinations am Center für Digitale Systeme der Freien Universität Berlin. Nach einer Lehre zum Bankkaufmann hat er an der Freien Universität Berlin und der Universität Potsdam Soziologie, Psychologie, Informatik und Statistik studiert.

Prüfungsdidaktik

Einführung in die Prüfungsdidaktik

Einleitende Bemerkungen

Ziel des vorliegenden Beitrags ist es, eine kurze Einführung in die Prüfungsdidaktik zu geben. Dazu soll zunächst eine Definition dieser Disziplin vorgenommen werden. Danach werden die Aufgaben und Ziele, der Geltungsbereich und einige grundlegende Prinzipien dieses neuen Faches behandelt; in diesem Rahmen wird es notwendig sein, Prüfungen zu klassifizieren. Hiernach wird eine Abgrenzung der Prüfungsdidaktik zu anderen (Teil-)Fächern vorgenommen.

In einem zweiten Schritt soll auf ausgewählte Aspekte der Durchführung mündlicher und schriftlicher Prüfungen eingegangen werden. Diese Ausführungen sind konkreten Prüfungen gewidmet und solchen Faktoren, die deren Gelingen bzw. Scheitern mitbedingen können.

Die hier vorgenommenen Beschreibungen und Reflexionen sind dabei fächerübergreifend angelegt - wie auch die Prüfungsdidaktik nicht auf ein bestimmtes Fach oder eine bestimmte Fachrichtung beschränkt ist. Nichtsdestotrotz ist sie bisher dominant auf den Fremdsprachenunterricht angewandt worden (vgl. Tinnefeld, 2002; 2013).

Die folgenden Ausführungen beziehen sich im Wesentlichen auf die Monographie „Dimensionen der Prüfungsdidaktik“ (Tinnefeld, 2013). Um die expliziten Verweise darauf nicht überborden zu lassen, wird auf die entsprechenden Stellen in unserer Monographie lediglich in Form einer allgemeinen Orientierung verwiesen. Kommen wir nun jedoch zu der Frage, welche Elemente die Prüfungsdidaktik konstituieren.

Konstituenten der Prüfungsdidaktik

Begriffsdefinition

Der Begriff *Prüfungsdidaktik* kann wie folgt definiert werden:

Unter Prüfungsdidaktik verstehen wir die Lehre vom wissenschaftlichen Prüfen – unter Berücksichtigung aller dort tätigen Individuen, also der Prüferinnen / Prüfer und der Prüflinge –, deren Ziel es ist, Prüfungen methodisch zu motivieren, zu objektivieren und ggf. zu standardisieren, und somit den ad hoc-Charakter, der Prüfungen vielfach anhaftet, so weit wie möglich zu minimieren. Prüfungsdidaktik ist per se fächerübergreifend und bezieht sich auf alle Wissenschaftsbereiche, in denen Prüfungen durchgeführt werden. Sie bezieht sich ebenso auf alle Institutionen, an denen geprüft wird: Universitäten und Hochschulen sowie alle Schulformen, vom Gymnasium bis hin zur Hauptschule (einschließlich der Sonder-Schulformen). Sie schließt die Prüfungen selbst und deren Vorbereitung durch den Prüfer ebenso ein wie die Vorbereitungen der Prüflinge auf sie – also auch den zu den Prüfungen hin-führenden Unterricht. Die Prüfungsdidaktik stellt somit eine gesamthafte, neue Wissenschaft dar (Tinnefeld, 2002, S. 6; 2013, S. 84).

Aus dieser Definition ergibt sich die Notwendigkeit der *Objektivierung*, der *Methodisierung* und gegebenenfalls der *Standardisierung* von Prüfungen. Die folgenden Reflexionen sind auf der Basis dieser Vorgaben zu sehen.

Aufgaben und Ziele der Prüfungsdidaktik

Die Aufgaben der Prüfungsdidaktik (vgl. Tinnefeld, 2013, S. 84 ff.) bestehen in einem ersten Ansatz darin, Prüfungen auf mehr oder minder optimale Art und Weise zu planen und umzusetzen. Dabei geht es zunächst darum, Prüfungen objektiv zu gestalten.

Als objektiv kann eine Prüfung dann bezeichnet werden, wenn das Ergebnis in weitgehender personaler Unabhängigkeit vom Prüfer zustande kommt (vgl. auch Jäger & Lissmann, 2000, S. 185-186). Im Rahmen der Objektivität von Prüfungen lassen sich die folgenden Untergruppen als die wichtigsten identifizieren:

- **Fachliche Objektivität:**
Hier geht es im Wesentlichen um die zu prüfenden fachlichen Inhalte und eine realistische Abgrenzung dieser im Hinblick auf die anvisierte Adressatengruppe.
- **Psychologische Objektivität:**
Die *psychologische Objektivität* bezieht sich zum einen auf den Prüfungstyp (mündliche / schriftliche Prüfung bzw. Einzel- / Gruppenprüfung) und zum

anderen auf die Frage, ob ein Prüfling sich in Stresssituationen leicht beeindrucken lässt und dadurch sein Leistungspotenzial nicht vollständig abrufen kann.

- Interaktionale Objektivität:

Bei der *interaktionalen Objektivität* handelt es sich um die Kommunikation zwischen Prüfer und Prüfling, wie sie sich auf mündliche, aber auch auf schriftliche Prüfungen bezieht. Es geht hier darum zu gewährleisten, dass in Prüfungen keine Missverständnisse zwischen beiden Interaktanten entstehen. Diese können u. a. dadurch entstehen, dass Prüfer und Prüfling sich nicht hinreichend gut kennen und sich aus diesem Grunde akademisch nicht hinreichend einschätzen können.

- Qualitative Objektivität:

Die *qualitative Objektivität* von Prüfungen bezieht sich auf die Kriterien *Validität* und *Reliabilität*. Das Qualitätskriterium *Reliabilität* (vgl. auch Jäger & Lissmann, 2000, S. 187-188) misst die Zuverlässigkeit von Prüfungen im Sinne ihrer Wiederholbarkeit, bei der die gleichen Ergebnisse erzielt werden sollten wie in der ursprünglichen Prüfung. Die Erfüllung dieses Kriteriums ist in der Prüfungspraxis recht schwer umzusetzen. *Validität*¹ bezieht sich im vorliegenden Zusammenhang auf die Gültigkeit von Prüfungen in dem Sinne, dass durch eine Prüfung exakt die Kenntnisse oder Fertigkeiten abgeprüft werden, die geprüft werden sollen. Wenn also beispielsweise in einer Prüfung praktisches Wissen getestet werden soll, kann dies nicht mit Hilfe von Multiple-Choice-Aufgaben geschehen, da diese hierfür einen ungeeigneten Aufgabentyp darstellen.

- Quantitative Objektivität:

Zu der qualitativen Objektivität von Prüfungen kommt deren *quantitative Objektivität*. Diese ist aus der Sicht der Prüflinge zum einen auf die Anzahl hintereinander erfolgreicher mündlicher oder schriftlicher Prüfungen bezogen und zum anderen auf die Länge schriftlicher Prüfungen und betrifft somit die Chancen der Prüflinge, das jeweils geforderte Pensum vollständig zu bearbeiten. Aus Prüfersicht bezieht sich die quantitative Objektivität auf die Arbeitsbelastung solcher Prüfer, die eine zu große Anzahl an Klausuren ebenso wenig mit hinreichender Qualität korrigieren können wie sie in der Lage sein werden, eine zu große Zahl mündlicher Prüfungen hintereinander in qualitativ befriedigender Art und Weise abzunehmen.

¹ Eine sehr zugängliche Definition des Begriffs *Validität* findet sich in Straub, Kempf & Werbik (2005, S. 573).

Vor diesem Hintergrund verfolgt die Prüfungsdidaktik das dringend bestehende Desiderat, die praktische Fertigkeit des Prüfens auf der Basis einer fundierten, ebenfalls im Rahmen der Prüfungsdidaktik zu erarbeitenden Theorie zu vermitteln. Hierfür sollte ein *Institut für Prüfungsdidaktik* geschaffen werden, in dem Prüfer beraten und im Sinne einer konstruktiven Qualitätskontrolle begleitet werden.

Geltungsbereich der Prüfungsdidaktik

Im Folgenden soll in aller gebotenen Kürze auf den Geltungsbereich der Prüfungsdidaktik (vgl. Tinnefeld, 2013, S. 104 ff.) eingegangen werden. Die Prüfungsdidaktik:

- ist in interner Abgrenzung in dem Sinne Bestandteil der wissenschaftlichen (Teil-)Fächer (vgl. hierzu auch Tinnefeld, 1993, S. 47 ff.; 2002, S. 13-14), dass Prüfungen ebenso in den Geisteswissenschaften (wie z. B. in der Anglistik oder der Romanistik) durchgeführt werden wie beispielsweise in den Naturwissenschaften (z. B. Biologie, Chemie), den Gesellschaftswissenschaften (z. B. Wirtschaftswissenschaften, Pädagogik) und den Technischen Wissenschaften (z. B. Elektrotechnik, Architektur), und
- hat – in entgegengesetzter Perspektive – diese Fächer als ihre eigenen Bestandteile: Prüfungsdidaktische Fragestellungen sind ebenso in den Geisteswissenschaften von zentralem Interesse wie in den Naturwissenschaften, den Gesellschaftswissenschaften und den Technischen Wissenschaften.

Aus dieser Verflechtung in beiden Richtungen kann eine *Relation mutualer Inklusion* abgeleitet werden.

In externer Abgrenzung kann die Prüfungsdidaktik als auf der jeweiligen Fachdidaktik einer gegebenen Wissenschaft aufbauend betrachtet werden. Dabei kann die jeweilige Fachdidaktik als *holistisch-generalisierend*, die Prüfungsdidaktik hingegen *partitiv-spezialisierend* definiert werden. In chronologischer ebenso wie in teleologischer Perspektive – also mit Blick auf ihre Zielgerichtetheit – ist die Fachdidaktik der Prüfungsdidaktik vorgelagert und auf diese ausgerichtet. Die Fachdidaktik kann somit als die Basis der Prüfungsdidaktik angesehen werden.

Im speziellen Falle von Sprachprüfungen ist die Testtheorie Teil der Prüfungsdidaktik. Dabei ist jeder standardisierte Test eine Prüfung, aber nicht jede Prüfung ein standardisierter Test. Zudem ist nicht jede Prüfung sinnvollerweise als standardisierter Test durchzuführen.²

² Vgl. hinsichtlich der Unterscheidung zwischen standardisierten und nicht-standardisierten Tests auch Lienert & Raatz (1998, S. 14) und zu derjenigen zwischen formellen und informellen Tests auch Dlaska & Krekeler (2009, S. 32).

Im Folgenden soll nun auf die wichtigsten Prinzipien der Prüfungsdidaktik eingegangen werden.

Prinzipien der Prüfungsdidaktik

Nachstehend (vgl. Tinnefeld, 2013, S. 121 ff.) seien wesentliche Merkmale professionell umgesetzter Prüfungen behandelt, wobei hier betont sei, dass die Planung und die Durchführung von Prüfungen einander bedingen. Prüfungspraxis und die dieser unterliegende Prüfungstheorie stehen somit in einem engen Verhältnis zueinander.

Die Prüfungsdidaktik bezieht sich dabei prinzipiell auf alle Institutionen und alle Wissenschaftsbereiche, in denen Prüfungen durchgeführt werden. Sie deckt die jeweilige Prüfung selbst, die Vorbereitung dieser und ebenso den auf sie vorbereitenden Unterricht ab und ist in diesem Sinne eine gesamthafte Wissenschaft.

Nun jedoch zu den einzelnen, für die Prüfungsdidaktik relevanten Prinzipien.

Sachliche Fundiertheit

Das Prinzip *Sachliche Fundiertheit* bezieht sich sowohl auf die fachwissenschaftliche wie auch die prüfungsdidaktische Ausrichtung von Prüfungen und ebenso auf schriftliche wie auf mündliche Prüfungen. Entscheidend für die Anwendung dieses Prinzips ist die Fähigkeit von Prüfern, Prüfungen zu antizipieren, also den möglichen Prüfungsverlauf - ggf. in unterschiedlichen Varianten - im Vorhinein mental durchzuspielen. Für die Ausbildung dieser Fähigkeit, die in aller Regel weniger auf Intuition als Erfahrung beruht, ist die umfangreiche Nachbereitung bereits abgehaltener Prüfungen von zentraler Bedeutung, wobei die Nachbereitung einer durchgeführten Prüfung als Teil der Vorbereitung der nächsten Prüfung angesehen werden kann.

Fachliches Aspirationsniveau

Das Prinzip *Fachliches Aspirationsniveau* bezieht sich auf die Notwendigkeit, einen hohen fachwissenschaftlichen Anspruch von Prüfungen bei gleichzeitiger prüfungsdidaktischer Orientierung zu gewährleisten. Eine generelle prüfungsdidaktische Orientierung von Prüfungen ist dabei nicht in dem Sinne misszuverstehen, dass diese eben durch diese prüfungsdidaktische Ausrichtung fachlich leichter würden. Angestrebt wird hier vielmehr, dass die Prüflinge durch die Orientierung von Prüfungen an der Prüfungsdidaktik während der und durch die Prüfung einen fachlichen Erkenntnisgewinn erzielen. In diesem Zusammenhang wird dringend empfohlen, von Prüflingen Transferleistungen zu fordern (vgl. Kapitel *Interaktion Prüfer / Prüfling* und *Detailfragen*).

Adressatenorientierung

Die Leitlinie, die im Rahmen des Prinzips *Adressatenorientierung* anzuwenden ist, besteht darin, die Prüflinge nur mit solchen Anforderungen zu konfrontieren, die sie auch erfüllen können. Eine Grundvoraussetzung dafür ist eine recht hohe Empathiefähigkeit des Prüfers. In praktischer Hinsicht geht es für Prüfer beispielsweise darum, Fragen in schriftlichen und mündlichen Prüfungen auf gut verständliche Art und Weise zu formulieren – eine simpel anmutende Forderung, die in der Praxis jedoch nicht entsprechend umgesetzt wird.

Valorisierung

Unter dem Begriff *Valorisierung* verstehen wir hier das Bestreben des Prüfers, seine Prüflinge in der Prüfungssituation zur Geltung kommen zu lassen. In diesem Sinne können und sollten Prüfer und Prüfling als Partner betrachtet werden. Die Prinzipien *Valorisierung* und *Adressatenorientierung* sind somit zwei Seiten der gleichen Medaille.

Fairness

Da Prüfer sich in aller Regel in einer weitgehend unangefochtenen und unanfechtbaren Position befinden, sollten sie hinsichtlich dieser Position und ihrer Funktion ein – möglichst seismographisches – Bewusstsein entwickeln, das ein potenziell missbräuchliches Verhalten möglichst verhindert. In mündlichen Prüfungen bedeutet dies für Prüfer beispielsweise, nicht in Monologe abzugleiten und nicht absichtlich Fragen hinsichtlich fachlicher Nischen zu stellen, um die Prüflinge in Schwierigkeiten zu bringen. Zudem sollten Prüfer alle möglichen Anstrengungen dahingehend unternehmen, jegliche Sympathie für und jegliche Antipathie gegen den jeweiligen Prüfling auszublenden.

Benotungsadäquatheit

Die Verbesserung der *Benotungsadäquatheit* von Prüfungen stellt einen hohen Anspruch dar. Zwar ist diese für standardisierte Prüfungen dank der dort zum Tragen kommenden Gütekriterien *Objektivität*, *Reliabilität* und *Validität* bereits weitgehend gewährleistet. Das generelle Problem liegt jedoch darin, dass die breite Mehrheit von Prüfungen keine standardisierten Prüfungen sind und diese Kriterien sich somit nicht auf sie anwenden lassen, so dass die Situation in nicht-standardisierten Prüfungen sich derzeit noch als durchaus unbefriedigend darstellt.

Diese Prinzipien stellen die Grundlage für unsere weiteren Ausführungen dar. Im Folgenden wird es nun darum gehen, einige Hinweise für die Durchführung zunächst mündlicher und danach schriftlicher Prüfungen zu geben.

Durchführung mündlicher Prüfungen

Allgemein sei zunächst betont, dass mündliche Prüfungen (vgl. Tinnefeld, 2013, S. 211 ff.) schwieriger planbar sind als schriftliche Prüfungen. Dieser Umstand ist wesentlich darin begründet, dass in mündlichen Prüfungen die Individualität des Prüfers einen noch stärkeren Einflussfaktor darstellt als in schriftlichen Prüfungen. Dies bedeutet in letzter Konsequenz, dass jeglicher Prüfer zu Selbstkritik fähig sein sollte, um mögliche Verhaltens- oder gar Charakterdefizite bestmöglich auszugleichen.

Für die Durchführung mündlicher Prüfungen sind im Wesentlichen die folgenden Aspekte im Blick zu behalten:

Das situative Umfeld

Die für das situative Umfeld mündlicher Prüfungen relevanten Faktoren lassen sich tabellarisch wie folgt darstellen:

Tabelle 1: *Situatives Umfeld mündlicher Prüfungen (vgl. Tinnefeld, 2013, S. 216)*

Prüfungsraum	Prüfer vs. Prüfling	Beisitzer	Zuhörer
<ul style="list-style-type: none">• Angst-reduzierende Atmosphäre• Angenehme Temperatur• Niedriger Lärmpegel	<ul style="list-style-type: none">• Angemessener räumlicher Abstand• Prüfertisch: geordnetes Arrangement	<ul style="list-style-type: none">• Möglichste Unauffälligkeit• Vermeidung von Beflissenheit beim Mitschreiben	<ul style="list-style-type: none">• Nur in Ausnahmefällen zuzulassen

Da diese Faktoren im Einzelnen aus sich heraus verständlich sind, erübrigt sich an dieser Stelle eine weitere Beschreibung (vgl. jedoch für mehr Details Tinnefeld, 2013, S. 213 ff.).

Gruppen- oder Einzelprüfung

Im Rahmen der grundsätzlichen Frage, ob mündliche Prüfungen als Gruppen- oder Einzelprüfung konzipiert werden sollten, empfehlen wir hier dringend die Vermeidung von Gruppenprüfungen und die ausschließliche Nutzung mündlicher Einzel-

prüfungen. Im Unterschied zu Gruppenprüfungen, in denen sich die Prüflinge nicht selten auf der Basis der Schwächen ihrer Mitkonkurrenten profilieren und eine Bewertungsorientierung an der Kohorte der zusammen geprüften Kandidaten vorgenommen wird, basieren die Ergebnisse von Einzelprüfungen auf der Leistungsfähigkeit des Individuums. Nur diese kann eine seriöse Referenzgröße sein.

Interaktion Prüfer / Prüfling

In Bezug auf die Interaktion zwischen Prüfer und Prüfling sind die folgenden Gesichtspunkte von Bedeutung:

- Geistige Präsenz des Prüfers:

Ein entscheidender Faktor für den Erfolg des Prüflings ist ein geistig präsender Prüfer, der sich auf jede mündliche Prüfung separat vorbereitet. Aufgrund dieser fundierten Vorbereitung kann er die einzelnen Prüfungsfragen logisch aufeinander aufbauen und sie zudem aus den Äußerungen des Prüflings ableiten. Er kann darüber hinaus den Schwierigkeitsgrad seiner Fragen für den Prüfling einschätzen und das Wissen des Prüflings in seiner ganzen Breite und Tiefe ausloten. Zudem entwickelt er mit jeder Prüfung mehr Prüfungskompetenz und steht somit jeglicher „Massenabfertigung“ in (hoch)schulischen Prüfungen entgegen. Auch wenn hier ein idealisiertes Bild beschrieben wird, sollte von jedem Prüfer ein entsprechendes Verhalten angestrebt werden.

- Prüferdisziplin:

Für Prüfer besteht die unbedingte Notwendigkeit, Selbstdisziplin zu üben – und zwar quantitativ und qualitativ. Die Notion *Quantitative Selbstdisziplin* bezieht sich dabei auf die Realisierung möglichst geringer Redeanteile durch den Prüfer, damit der Prüfling entsprechend zu Wort kommen kann. Die Notion *Qualitative Selbstdisziplin* bezieht sich auf ein Bewusstsein des Prüfers hinsichtlich des vom Prüfling realistisch erreichbaren Leistungsniveaus. Eine fruchtbare Strategie der Gesprächsführung kann darin bestehen, den Prüfling mit jeder Frage zu einem etwas höheren fachlichen bzw. akademischen Niveau zu führen.

In diesem Zusammenhang benötigt der Prüfer – neben einer hohen fachlichen Kompetenz – eine hohe Kommunikationsfähigkeit und eine ausgeprägte Fähigkeit zur Empathie.

- Vermeidung von Machtspielen:

In mündlichen Prüfungen besteht für Prüfer die generelle Versuchung, die Situation beherrschen zu wollen und sich als alleinige Autorität darzustellen. Prüfer, die so handeln, missbrauchen ihre Rolle und üben – bewusst

oder unbewusst – Macht aus. Eine solche Zurschaustellung von Macht wirkt sich potenziell negativ auf den Prüfungserfolg aus (vgl. hierzu auch Buchwald, 2007, S. 162), ist daher prüfungsdidaktisch kontraproduktiv und somit unbedingt zu vermeiden.

- Schaffung von Angstfreiheit:

Die oberste Priorität muss für Prüfer in prüfungsdidaktischer Perspektive darin liegen, eine Atmosphäre der Angstfreiheit zu schaffen:³ Nur diese ermöglicht dem Prüfling die Erbringung optimaler Leistungen und die vollständige Abrufung seines persönlichen Leistungspotenzials. Der Prüfer muss also sein Möglichstes dafür tun, Ruhe auszustrahlen und die Prüfung schrittweise inhaltlich entwickeln, um auch in thematischer Hinsicht Steigtigkeit zu realisieren. Zudem sollte er den Prüfling zu Beginn und während der Prüfung aufgrund seiner vorherigen und ebenso seiner aktuellen Prüfungsleistungen loben, um ihn auf diese Weise mental zu stärken.

- Realisierung der mündlichen Prüfung als Gespräch:

Mündliche Prüfungen sollten keine Frage-Antwort-Spiele, sondern vielmehr als (akademische) Gespräche konzipiert sein. Ein dafür höchst geeigneter Fragetyp ist die Transferfrage. Zur Illustration sei das folgende Beispiel angeführt:

„Bewerten Sie den Einfluss der Textlinguistik auf die Entwicklung der Linguistik im Allgemeinen.“

Eine solche Art der Fragestellung ermöglicht ein Hinausgehen über reines Buchwissen und eine Bezeugung eines tieferen Verständnisses von Sachzusammenhängen. Transferfragen wie diese schließen Fachwissen mit ein. Zudem sind sie bei brüchigem Faktenwissen kaum schlüssig zu beantworten und somit eine verlässliche Frageform der Leistungsermittlung. Sie stellen somit für mündliche Prüfungen einen geradezu idealen Fragetyp dar.

Kommen wir nun zu Aspekten der Durchführung schriftlicher Prüfungen.

³ Vgl. hinsichtlich der negativen Auswirkungen von Angst und Stress auf die Leistungen von Prüflingen auch Buchwald (2002).

Durchführung schriftlicher Prüfungen

Im Rahmen schriftlicher Prüfungen sollen im Folgenden – in Anlehnung an Tinnefeld (2013, S. 343 ff.) – verschiedene Aufgabentypen dargestellt und im Sinne ihrer prüfungsdidaktischen Vor- und Nachteile analysiert werden.

True-False-Statements

Im Rahmen des Aufgabentyps *True-False-Statement*⁴ wird hier für dessen Variante mit einer Korrekturmöglichkeit – also einem aktiven Teil, der das rein rezeptive Ankreuzen der beiden Antwortalternativen relativiert – plädiert. Ein Beispiel möge dies verdeutlichen:

Die Linguistik ist eine reine Geisteswissenschaft.

Wahr Falsch

Mögliche Korrektur: _____

Die prüfungsdidaktischen Vorteile dieses Aufgabentyps liegen in der relativ einfachen Erstellung der einzelnen Items, deren rascher Korrektur und der klaren Aufgabenstruktur für die Prüflinge. Zudem liegen sie in der hier gezeigten Kombination rezeptiver und produktiver Elemente und in der Möglichkeit der Erfassung von Global- und Detailinformationen.

Nachteile von True-False-Statements sind im Grunde nur dann gegeben, wenn deren Variante ohne jegliche Korrekturmöglichkeit gewählt wird, von der wir hier jedoch absehen.

Somit überwiegen die Vorteile dieses Aufgabentyps, der für die Abprüfung von Verstehensleistungen hier als wertvoll eingestuft wird.

Multiple-Choice-Aufgaben

Multiple-Choice-Aufgaben stellen einen weithin bekannten Aufgabentyp dar. Das folgende Beispiel aus der Linguistik möge ihrer Veranschaulichung dienen:

Die Verwendung des Passivs hat Auswirkungen auf:

a. *die Satzperspektive*

b. *die reine Satzbedeutung*

⁴ In diesem Zusammenhang sei auf die Empfehlungen von Landsberger (2015a) zu diesem Aufgabentyp verwiesen.

- c. die Textkonstitution
- d. kein Satzelement

[Korrekte Antwort: (a) und (c)]

Die Vorteile von Multiple-Choice-Aufgaben liegen in der Möglichkeit der Bereitstellung einer oder – wie im Beispiel veranschaulicht und hier dringend empfohlen – mehrerer richtiger Antwortalternativen, ihrer relativ raschen Erstellung und Korrektur und ihrem hohen Bekanntheitsgrad bei Prüflingen.

Die Nachteile dieses Ausgabentyps sind in seiner lediglich scheinbaren Objektivität, dem generellen Problem der Glaubhaftigkeit der Distraktoren, der ausschließlich rezeptiven Bearbeitung der Aufgaben und somit der Erbringung von Prüfungsleistungen im Rahmen eines reinen Wiedererkennens von Informationen zu sehen.

Unter der Bedingung, dass jeweils mehr als eine korrekte Lösung vorgesehen wird, können Multiple-Choice-Aufgaben als durchaus empfehlenswerter Aufgabentyp angesehen werden.

Detailfragen

Detailfragen stellen einen Fragetyp dar, mit dem die Kenntnis von für eine gegebene Wissenschaft relevanten Teilaspekten geprüft werden kann. Das folgende Beispiel aus dem Bereich der Linguistik soll diesen Aufgabentyp verdeutlichen. Hierbei wird eine sachbezogene Detailfrage gestellt und in der Klammer eine Möglichkeit angedeutet, diese in eine Transferfrage zu transformieren:

Was ist ein Phonem und was ist seine distinktive Funktion? (Geben Sie Beispiele aus einer von Ihnen gewählten romanischen Sprache und analysieren Sie diese in funktionaler Perspektive.)

Die Vorteile des Aufgabentyps *Detailfrage* liegen in seinem unmittelbaren Verständnis durch die Prüflinge, seiner Anwendbarkeit auf alle denkbaren Informationstypen – vorausgesetzt, es handelt sich um Teilaspekte bzw. ausgrenzbare Einzelheiten –, darin, dass er ein produktives Prüfungsformat darstellt – die Prüflinge also selbst Antworten formulieren müssen – sowie in seinen Möglichkeiten zur Abprüfung exakten Sachwissens.

Seine Nachteile sind in der potenziellen Willkür der abgeprüften Wissensbestandteile zu sehen – es könnte ja durchaus dieses und nicht jenes Detail abgefragt werden – und in der *per definitionem* geringen Stoffabdeckung dieses Aufgabentyps.

Insgesamt ist der Aufgabentyp *Detailfrage* – unter der Voraussetzung, dass er in Verbindung mit Globalfragen verwendet wird – als empfehlenswert einzustufen,

zumal er jederzeit Erweiterungen für Transferoptionen bietet (wie im Beispiel in der Klammer verdeutlicht).

Globalfragen

Auch für den Aufgabentyp *Globalfrage* sei zunächst ein Beispiel gegeben:

Beschreiben (und analysieren) Sie kurz die wichtigsten Entwicklungen der Linguistik in der zweiten Hälfte des 20. Jahrhunderts.

Die Vorteile dieses Aufgabentyps sind in seinen Möglichkeiten der Abprüfung eines allgemeinen thematischen Überblicks über ein gegebenes (Teil-)Fach zu sehen, wobei auch hier in sehr zugänglicher Form Transferanteile eingebaut werden können, was im Beispiel wiederum innerhalb der Klammer umgesetzt worden ist. Zudem ist es mit Hilfe dieses Aufgabentyps möglich, ein breites Themenspektrum mittels einer begrenzten Anzahl von Fragen abzudecken.

Ein möglicher Nachteil dieses Aufgabentyps ist es, dass Grenzziehungen zwischen diesem sowie *Detailfragen* und *Zusammenfassungen* nicht immer eindeutig möglich sind. In Verbindung mit diesen stellen *Globalfragen* jedoch einen wertvollen Aufgabentyp dar.

Zusammenfassungen

Für den Aufgabentyp *Zusammenfassung* sei hier ein Beispiel gegeben, das auf der Beigabe eines – gegebenenfalls adaptierten – wissenschaftlichen Textes einer Länge von ein bis zwei Druckseiten basiert:

Verfassen Sie ein Abstract zu dem Ihnen vorliegenden Fachartikel über die Textfunktionen des Passivs im Französischen.

Die Vorteile dieses Aufgabentyps sind sein potenziell holistischer Charakter, der die Erfassung der großen Linien eines Faches in gesteuerter Form ermöglicht, und die weitgehend problemlose Erstellung der jeweiligen Arbeitsinstruktion.

Seine Nachteile liegen in der aufwändigen Korrektur und der weitgehenden Unmöglichkeit der eindeutigen Festlegung klarer und objektivierbarer Bewertungskriterien.

Trotz dieser Nachteile kann die *Zusammenfassung* als ein – besonders im akademischen Bereich – hochrelevanter Aufgabentyp angesehen werden, der die Abprüfung der Abstraktionsfähigkeiten der Prüflinge auf der Basis gegebener Informationen unter Einschluss von deren gesamtem Fachwissen ermöglicht.

Fazit

Zum Abschluss unserer Reflexionen lässt sich festhalten, dass es sich bei den hier beschriebenen Prüfungssituationen um solche von hochkomplexem Charakter handelt, die von Seiten des Prüfers hohe Flexibilität und hohe Sensibilität, Fähigkeit zur Empathie gepaart mit einem hohen fachlichen Niveau und eine eingehende Vorbereitung jeder einzelnen Prüfung voraussetzen. Hinzu kommen die Notwendigkeit zu regelmäßiger Hinterfragung der eigenen Prüfertätigkeit und eine umfassende Informiertheit über existierende Aufgabentypen und deren konkrete Anwendung in mündlichen und schriftlichen Prüfungen.

Abschließend sei der Hoffnung Ausdruck verliehen, dass in Zukunft immer mehr Prüfer auf der Basis von Darstellungen wie der vorliegenden ihre eigene Prüfungspraxis optimieren und Prüfungen immer prüfungsdidaktischer gestalten werden.

Referenzen

- Buchwald, P. (2002). *Dyadisches Coping in mündlichen Prüfungen*. Göttingen: Hogrefe.
- Buchwald, P. (2007). Macht in Prüfungen: Eine ressourcenorientierte Analyse. In M. Göhlich, E. König & C. Schwarzer (Hrsg.), *Beratung, Macht und organisationales Lernen* (S. 161-170). Wiesbaden: Verlag für Sozialwissenschaften.
- Dlaska, A. & Krekeler, Ch. (2009). *Sprachtests: Leistungsbeurteilungen im Fremdsprachenunterricht evaluieren und verbessern*. Baltmannsweiler: Schneider Hohengehren.
- Jäger, R. S. & Lissmann, U. (2000). *Von der Beobachtung zur Notengebung: Ein Lehrbuch - Diagnostik und Benotung in der Aus-, Fort- und Weiterbildung*. Landau: Verlag Empirische Pädagogik.
- Landsberger, J. (2015a). Study guides and strategies: True/false tests. Abgerufen unter <http://www.studygs.net/tsttak2.htm> [01.05.2015]
- Landsberger, J. (2015b). Study guides and strategies: Multiple choice tests. Abgerufen unter <http://www.studygs.net/tsttak3.htm> [01.05.2015]
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Straub, J., Kempf, W. & Werbik, H. (2005). *Psychologie: Eine Einführung*. München: Deutscher Taschenbuch Verlag.
- Tinnefeld, T. (1993). *Die Syntax des Journal officiel': Eine Analyse der Fachsprache des Rechts und der Verwaltung im Gegenwartsfranzösischen*. Bochum: AKS.

Tinnefeld, T. (2002). *Prüfungsdidaktik: Zur Fundierung einer neuen wissenschaftlichen Disziplin – am Beispiel der modernen Fremdsprachen*. Aachen: Shaker.

Tinnefeld, T. (2005). Mündliche Prüfungen im Hochschulstudium: Möglichkeiten der Optimierung. In M. Ó Dúill, R. Zahn & K. D. C. Höppner (Hrsg.), *Zusammenarbeiten: Eine Festschrift für Bernd Voss* (S. 421-436). Bochum: AKS.

Tinnefeld, T. (2013). *Dimensionen der Prüfungsdidaktik: Analysen und Reflexionen zur Leistungsbewertung in den modernen Fremdsprachen*. Saarbrücken: Hochschule für Technik und Wirtschaft des Saarlandes.

Abgerufen unter <https://drive.google.com/file/d/oB-2HUXwM3-xiYXhXQndnSTVCcDg/edit?usp=sharing> [10.09.2015]

Vita

Seit 2008 W3-Professor für Angewandte Sprachen an der Hochschule für Technik und Wirtschaft des Saarlandes. Zahlreiche Publikationen in den Forschungsschwerpunkten *Linguistik* (Fachsprachenforschung, Metalexikographie, Grammatikographie und Schreibforschung) und *Fremdsprachendidaktik* (Fremdsprachenvermittlung, Prüfungsdidaktik und Interkulturelles Lernen). Herausgeber der internationalen Zeitschrift *Journal of Linguistics and Language Teaching*. Leiter der *Saarbrücker Fremdsprachentagungen*.

Michael Niedeggen
Freie Universität Berlin

Bulimielernen verhindern

Ein gravierendes Problem in den Bachelorstudiengängen ist die Fokussierung auf die schriftlichen Testate am Ende jedes Semesters. Durch die Konzentration der Prüfungen auf ein enges Zeitfenster wird das Lernverhalten der Studierenden stark negativ beeinflusst, da eine kontinuierliche Auseinandersetzung mit den Lerninhalten nicht stattfindet. Ich möchte ein alternatives Prüfungsmodell vorstellen, welches wir für das Modul *Biopsychologie* im ersten Studienjahr im Studiengang *Psychologie* eingeführt haben. Durch den Einsatz von Teilprüfungen wird der Prüfungsplan entzerrt und das Annahmeverhalten der Studierenden positiv beeinflusst. Die Umstellung der Prüfungen auf eine E-Examination-Form erleichtert nicht nur die Korrektur und beschleunigt damit die zeitnahe Rückmeldung an die Studierenden, sondern hat auch einen Effekt auf die Prüfungsleistung.¹

¹ Niedeggen, M. (2015). *Bulimielernen verhindern: Vortrag auf der Tagung "GML² 2015: Grundfragen Multimedialen Lehrens und Lernens – E-Examinations: Chances and Challenges."*
Abgerufen unter <http://www.gml-2015.de/tagungsprogramm/niedeggen/> [30.06.2015]

Sebastian Schubert
Charité – Universitätsmedizin Berlin

Der Progress Test Medizin: Eine multizentrische Kooperation zur Erfassung individueller Lernverläufe im Medizinstudium

In diesem Beitrag werden das Konzept des Progress Testings, dessen Umsetzung im deutschsprachigen Progress Test Medizin (PTM) und die Chancen und Herausforderungen der computergestützten Durchführung dargestellt.

Progress Testing ist die longitudinale Messung des Lernfortschritts in einem Lernprozess. Über die Dauer des Lernprozesses wird mit inhaltsäquivalenten Tests wiederholt das Wissen erfasst, das die Lernenden am Ende des Lernprozesses erworben haben sollen. Damit kann für jeden Lernenden eine individuelle Lernfortschrittentwicklung abgebildet und mit anderen Lernenden verglichen werden.

Der Progress Test Medizin (PTM) ist ein Kooperationsprojekt von 14 deutschen und drei österreichischen medizinischen Fakultäten. Er wird seit 1999 von der Arbeitsgruppe Progress Test Medizin entwickelt. Jeder Test besteht aus 200 Multiple-Choice-Fragen, die das Wissen abbilden, das Medizinstudierende am Ende ihres Studiums erworben haben sollen. Er wird zweimal pro Jahr angeboten und ist an allen kooperierenden Fakultäten formativ. Im Wintersemester 2014 / 15 nahmen ca. 10.000 Studierende teil.

Sechs Kooperationspartner führen den PTM computergestützt durch. In Berlin nutzt die Charité dazu das E-Examination Center (EEC) der Freien Universität. Vorteile der computergestützten Durchführung sind der Wegfall von Testunterlagen auf Papier und die sofortige Rückmeldung an die Teilnehmer. Andere Vorteile der computergestützten Durchführung, v. a. die Verwendung von Fotos, Audios und Videos, sind für den PTM zur Zeit nicht nutzbar, da die Mehrheit der Kooperationspartner den PTM noch auf Papier durchführt. Nachteile der computergestützten Durchführung sind die Kosten und die begrenzte Anzahl der Plätze im EEC.¹

¹ Schubert, S. (2015). *Der Progress Test Medizin – eine multizentrische Kooperation zur Erfassung individueller Lernverläufe im Medizinstudium: Vortrag auf der Tagung "GML² 2015: Grundfragen Multimedialer Lehrens und Lernens – E-Examinations: Chances and Challenges.*" Abgerufen unter <http://www.gml-2015.de/tagungsprogramm/schubert/> [30.06.2015]

Referenzen

Progress Test Medizin. Abgerufen unter <http://ptm.charite.de/> [28.05.2015]

Progress Testing. Abgerufen unter http://en.wikipedia.org/wiki/Progress_testing [28.05.2015]

Sicherheit

Prüfungsrechtliche Rahmenbedingungen für elektronische (Präsenz-)Prüfungen

Einleitung

Das herkömmliche Prüfungsrecht ist auf die „klassischen“ Formen der mündlichen und schriftlichen Prüfungen zugeschnitten. Rechtsprechung und Literatur haben sich bislang mit den rechtlichen Rahmenbedingungen für elektronische Prüfungen nur am Rande beschäftigt. Der nachfolgende Beitrag wird die von den Prüfungsbehörden zu beachtenden juristischen Besonderheiten bei elektronischen Prüfungen erläutern. Dabei sollen die Frage nach den an eine hinreichende Rechtsgrundlage zu stellenden Anforderungen sowie die Darstellung des einschlägigen Prüfungsverfahrensrechts im Mittelpunkt stehen.

Begrifflichkeit

Was ist unter einer elektronischen Präsenzprüfung im rechtlichen Sinne zu verstehen?

Begrifflich ist zunächst zu klären, was unter einer elektronischen Präsenzprüfung im rechtlichen Sinne zu verstehen ist. Auszugehen ist dabei von der im Prüfungsrecht allseits anerkannten klassischen Unterteilung in Aufsichtsarbeiten und Hausarbeiten. Bei Präsenzprüfungen im hiesigen Sinne handelt es sich um Prüfungen, bei denen die Prüfungsfragen an einem Computerbildschirm angezeigt werden und die Antworten sogleich an diesem „stationären“ Computer eingegeben werden, der Computer mithin nicht bloß als „Schreibmaschinenersatz“ genutzt wird. Es handelt sich also insbesondere um Aufsichtsarbeiten – oftmals im Antwort-Wahl-Verfahren –, die in elektronischer Form stattfinden. In der Sache zielt dieses Verfahren darauf ab, dass die auf einem Bildschirm angezeigten Prüfungsfragen ausschließlich durch das Markieren der vom Anwendungsprogramm vorgegebenen Antwortfelder mit einem Eingabegerät beantwortet werden und die Fragen und

Antworten ausschließlich als digitale Informationen auf einem Speichermedium verbleiben.¹

Abgrenzung zur elektronischen Übermittlung von Prüfungsleistungen

Abzugrenzen davon sind Konstellationen, in denen eine vom Prüfling angefertigte Prüfungsleistung – in der Regel dürfte es sich hierbei um eine Hausarbeit handeln – in elektronischer Form der Prüfungsbehörde übermittelt wird. Hier sehen die herkömmlichen Prüfungsordnungen für häusliche Arbeiten zumeist die Schriftform vor. Damit gemeint ist die Vorlage eines Schriftstücks mit persönlicher Unterschrift. Die Fertigung und Übersendung in elektronischer Form² ist nicht in diesem Sinne „schriftlich.“³ Wenngleich also zwischen Schriftform und elektronischer Kommunikation grundsätzlich zu unterscheiden ist, gestattet § 3a des Verwaltungsverfahrensgesetzes – VwVfG – unter näher bezeichneten Voraussetzungen, die durch Rechtsvorschrift angeordnete Schriftform durch die elektronische Kommunikation zu ersetzen, und zwar auch für den Bereich des Prüfungswesens (§ 2 Abs. 3 Nr. 2 VwVfG).⁴ Durch die nach § 3a Abs. 2 VwVfG zwingend erforderliche Verwendung einer qualifizierten elektronischen Signatur soll erreicht werden, dass den Sicherungsfunktionen der an sich gebotenen Schriftform entsprochen wird.⁵ Ob und wieweit diese gesetzliche Möglichkeit des Ersatzes der vorgesehenen Schriftform für häusliche oder praktische Prüfungsarbeiten realisierbar ist, muss die Praxis erweisen.⁶ Die Prüfungsordnung kann dies ausdrücklich ausschließen (§ 3a Abs. 2 Satz 1, 2. Halbs. VwVfG). Sieht hingegen die Prüfungsordnung keine zwingende Schriftform vor, sondern regelt selbst die Online-Übermittlung von zuvor vom Prüfling am eigenen Computer erstellten Prüfungsleistungen (per E-Mail oder auf ähnlichem Wege), so steht einer elektronischen Übermittlung auch ohne qualifizierte elektronische Signatur nichts im Wege, wobei dann allerdings die Authentizität

¹ VG Hannover, Beschluss vom 10. Dezember 2008 – 6 B 5583/08 – juris Rdn. 32, sowie Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 ff. Anders und insoweit zweifelhaft OVG Koblenz, Beschluss vom 19. Januar 2009 – 10 B 11244/08 – juris Rdn. 9, wonach die Schriftlichkeit auch dann noch gewahrt sein soll, wenn auf im Computer gestellte Fragen ebenfalls dort vorgegebene Antwortalternativen per Mausclick angekreuzt werden müssen.

² Ein Überblick zur Prüfungsform der E-Klausur findet sich bei Vogt / Schneider, E-Klausuren an Hochschulen: Didaktik – Technik – Systeme – Recht – Praxis, Gießen 2009. (<http://geb.uni-giessen.de/geb/volltexte/2009/6890>).

³ VG Hannover, Beschluss vom 10. Dezember 2008 – 6 B 5583/08 – juris Rdn. 32.

⁴ S. dazu auch Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (314).

⁵ Dazu im Einzelnen: Kopp / Ramsauer, Verwaltungsverfahrensgesetz, 14. Aufl. 2013, § 3a Rdn. 14 a. und 17.

⁶ S. zu technischen Problemen und Manipulationsgefahr bei elektronischen Prüfungen: Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (317 ff.).

tät der Prüfungsleistung durch zusätzliche Maßnahmen (etwa schriftliche Versicherung des Prüflings, dass er die Prüfungsleistung selbst erbracht hat) sicherzustellen ist.⁷

Rechtlicher Überbau – der Gesetzesvorbehalt

Es ist heute allgemein anerkannt, dass Prüfungen eine gesetzesförmige Rechtsgrundlage benötigen. Der Vorbehalt des Gesetzes gilt auch für das Prüfungswesen.⁸ Leitentscheidungen des Gesetzgebers sind in diesem Rechtsbereich insbesondere wegen der Grundrechtsrelevanz von Prüfungsentscheidungen erforderlich. Da (negative) Prüfungsentscheidungen in aller Regel die grundgesetzlichen Freiheiten der Berufswahl oder der Wahl der Ausbildungsstätte einschränken, ist der Schutzbereich der Berufsfreiheit (Art. 12 Abs. 1 des Grundgesetzes – GG) durch sie berührt. Allerdings ist es in der Rechtsprechung nicht abschließend geklärt, welche Anforderungen der Gesetzesvorbehalt an die Regelung des Prüfungswesens stellt.⁹ Insbesondere die Frage, welche Leitentscheidungen der parlamentarische Gesetzgeber selbst treffen muss und welche Entscheidungen dem Verwaltungs- oder Satzungsgeber, also etwa der Universität als Autor der Prüfungsordnungen, überlassen bleiben können, ist nicht eindeutig zu beantworten. Gleichwohl ist eine tendenziell großzügige Haltung der Rechtsprechung zu erkennen, was die eigenverantwortlichen Regelungskompetenzen der Exekutive betrifft.¹⁰

Dass wesentliche Merkmale der Ausgestaltung des Prüfungsverlaufs normativ zu regeln sind, haben Rechtsprechung und Literatur mehrfach zum Ausdruck gebracht.¹¹ „Wesentlich“ in diesem Sinne ist nicht zuletzt, in welcher Form die Prüfung stattfindet. Traditionell wurde durchweg zwischen schriftlichen, mündlichen oder praktischen Prüfungen unterschieden. Eine ausschließlich *elektronische Präsenzprüfung*, bei der die Prüfungsfragen im Computerbildschirm angezeigt werden und die Antworten sogleich an diesem „stationären“ Computer eingegeben werden, der Computer also nicht bloß als „Schreibmaschinenersatz“ genutzt wird, ist *nicht* als eine Art *schriftliche Prüfung* zu verstehen, sondern muss angesichts ihrer technischen Besonderheiten und Unsicherheiten hinsichtlich der Authentizität des

⁷ Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 428.

⁸ Grundlegend ist der Beschluss des BVerfG vom 20. Oktober 1981 – 1 BvR 640/80, BVerfGE 58, 257 = DVBl. 1982, 401 = NJW 1982, 921; fortgesetzt und präzisiert durch Beschlüsse vom 14. März 1989 – 1 BvR 1033/82, 174/84, BVerfGE 80, 1 = NVwZ 1989, 850 und vom 17. April 1991 – 1 BvR 419/81, 213/83, BVerfGE 84, 34 = NJW 1991, 2005 und – 1 BvR 1529/84, 138/87, BVerfGE 84, 59 = NJW 1991, 2008. S. auch Becker, Parlamentsvorbehalt im Prüfungsrecht, NJW 1990, 274.

⁹ Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 21.

¹⁰ Dazu näher Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 22 f.

¹¹ Dazu insbesondere BVerfG, Beschluss vom 14. März 1989 – 1 BvR 1033/82, 174/84, BVerfGE 80, 1 = NVwZ 1989, 850 = DVBl. 1989, 814.

Urhebers als eine andersartige Form der Leistungsermittlung angesehen werden;¹² sie benötigt daher eine *ausdrückliche normative Grundlage*.¹³ Sofern eine derartige normative Grundlage – etwa in der Prüfungsordnung – vorliegt, bestehen jedoch im Hinblick auf § 3a Abs. 1 VwVfG keine Bedenken dagegen, dass elektronische Präsenzprüfungen durchgeführt werden. Nicht ausreichend für die Durchführung elektronischer Prüfungen ist hingegen die bloße Erwähnung einer „schriftlichen“ Prüfung in der entsprechenden normativen Regelung.¹⁴

Rechtliche Anforderungen an die Ausgestaltung von elektronischen Prüfungsverfahren

Allgemeine Anforderungen an das Prüfungsverfahren

Grundsätzlich sind die allgemeinen prüfungsrechtlichen Grundsätze, die sich insbesondere aus dem Gebot der Chancengleichheit (Art. 3 Abs. 1 GG) und den Notwendigkeiten, die ein Eingriff in die Berufsfreiheit (Art. 12 Abs. 1 GG) bedingt, ableiten, auch auf elektronische Prüfungen anzuwenden.

Wie bei anderen Prüfungen greifen auch bei elektronischen Prüfungen die *allgemeinen Rechte und Pflichten der Prüflinge*. Sie haben einen Anspruch auf Zulassung zur Prüfung und auf Durchführung eines fairen Prüfungsverfahrens. Das beinhaltet die rechtzeitige Ladung zur Prüfung und ebenso, dass die dafür zuständigen Prüfer die Prüfungsaufgaben erstellen und die Prüfungsleistungen im Anschluss bewerten. Dabei ist zu fordern, dass die Erstellung und Bewertung der Prüfungsleistung durch unvoreingenommene, also nicht befangene Prüfer erfolgt, die zudem über die erforderliche fachliche Qualifikation verfügen müssen. Der abgeprüfte Prüfungs-

¹² Zweifelhaft OVG Koblenz, Beschluss vom 19. Januar 2009 – 10 B 11244/08 – juris Rdn. 9, wonach die Schriftlichkeit auch dann noch gewahrt sein soll, wenn der Prüfling die Aufgabenstellung im Computer wahrnehmen und dann per Mausklick ankreuzen muss, welche der ebenfalls im Computer ersichtlichen Antworten richtig ist.

¹³ VG Hannover, Beschluss vom 10. Dezember 2008 – 6 B 5583/08 – juris Rdn. 32, wonach eine abweichende rechtliche Einordnung einer elektronischen Prüfung als schriftliche Prüfung allerdings ggf. dann in Betracht komme, wenn festgelegt sei, dass die Aufzeichnungen des Anwendungsprogramms über die eingegebenen Prüfungsfragen im Zusammenhang mit den Markierungen der Antworten und mit Hilfe eines ausreichend sicheren technischen Nachweises ihrer Authentizität ausgedruckt und bei den Prüfungsunterlagen aufbewahrt werden und dort eingesehen werden können. Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (314 f.).

¹⁴ VG Hannover, Beschluss vom 10. Dezember 2008 – 6 B 5583/08 – juris Rdn. 32, sowie Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 ff. Anders und insoweit zweifelhaft OVG Koblenz, Beschluss vom 19. Januar 2009 – 10 B 11244/08 – juris Rdn. 9.

stoff muss schließlich zulässig sein und darf nicht über das Ziel der Prüfung, nämlich in der Regel eine bestimmte Berufsqualifikation festzustellen, hinausschießen.

Ferner haben die Prüflinge einen Anspruch darauf, die Prüfung ohne störende äußere Einflüsse wie etwa Lärm oder Kälte abzuleisten. Auch haben sie nach den ganz generell geltenden Grundsätzen die Möglichkeit von der Prüfung zurückzutreten, wenn sie krank sind, oder unter bestimmten Voraussetzungen¹⁵ einen Nachteilsausgleich zu verlangen, wenn eine Behinderung vorliegt. Die Prüflinge haben selbstverständlich auch die Pflicht, Täuschungen zu unterlassen und die Leistung persönlich zu erbringen. Nach erfolgter Prüfung haben sie zur Wahrung und ggf. gerichtlicher Durchsetzung ihrer Rechte einen Anspruch auf Akteneinsicht.

Besonderheiten bei elektronischen Prüfungen

Von besonderem Interesse ist im hiesigen Kontext, welche verfahrensrechtlichen Besonderheiten sich vor, bei und nach elektronischen Prüfungen stellen.

Im Vorfeld der Prüfung

Bei der *Anmeldung* zur Prüfung ist die *Authentifizierung* des Prüflings sicherzustellen. Soll die Anmeldung auf elektronischem Wege erfolgen, kann nicht auf die herkömmlichen Legitimationsmittel wie Ausweispapier und Unterschrift abgestellt werden. Sieht die Prüfungsordnung grundsätzlich die Schriftform für die Anmeldung vor, so ist dann gemäß § 3a Abs. 2 VwVfG eine qualifizierte elektronische Signatur zur Anmeldung erforderlich. Ohne Schriftformerfordernis ist zwar eine elektronische Signatur nicht erforderlich, allerdings ist dann eine Authentifizierung des sich anmeldenden Prüflings auf andere Weise – etwa durch ein PIN / TAN-Verfahren – sicherzustellen.¹⁶

Soll für jeden Prüfling eine *individualisierte Klausur* bereitgestellt werden, so ist im Hinblick auf das Chancengleichheitsgebot ein *einheitlicher Schwierigkeitsgrad* sicherzustellen; dies kann etwa derart geschehen, dass alle Prüflinge zwar dieselben Fragen gestellt bekommen, allerdings in unterschiedlicher Reihenfolge.¹⁷ Werden hingegen den einzelnen Prüflingen unterschiedliche Fragen aus einem standardisierten Fragenkatalog zur Beantwortung vorgelegt, muss zum einen sichergestellt sein, dass die jeweiligen individuellen Zusammenstellungen der Fragen den gleichen Schwierigkeitsgrad haben. Außerdem erscheint es vor dem Hintergrund, dass der Prüfer selbst die Prüfungsleistung eigenverantwortlich bewerten muss, nicht

¹⁵ Dazu Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 258 ff.

¹⁶ Vgl. hierzu http://ep.elan-ev.de/wiki/Rechtsfragen#cite_ref-16, Stichwort: „An- und Abmeldung“; Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 33 f. sowie VG Saarlouis, Beschluss vom 23. Juli 1998 – 1 F 73/98 – NJW 1998, 3221.

¹⁷ Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 38.

unproblematisch, wenn die Zusammenstellung der individuellen Klausurfragenkataloge automatisiert durch einen Computer erfolgen sollte. Allerdings kann man sich auf den Standpunkt stellen, dass bereits die bloße Standardisierung des Fragenkatalogs durch den Prüfer insoweit ausreicht.¹⁸

Aus der der Prüfungsbehörde obliegenden *Informationspflicht* leitet sich der Anspruch der Prüflinge darauf ab, bereits im Vorfeld der elektronischen Prüfung mit den wesentlichen technischen Besonderheiten – etwa im Rahmen von *Übungsklausuren* – vertraut gemacht zu werden, damit sichergestellt ist, dass sie über die erforderlichen technischen Kenntnisse verfügen, um die Prüfungsleistung ohne Zeitverlust aufgrund technischer Anwendungsprobleme durchführen zu können.¹⁹

Während der Prüfung

Das Gleichbehandlungsprinzip ist auch während der Prüfung zu wahren; insbesondere muss die *gleiche Ausstattung* der für die Prüfung zur Verfügung gestellten Computer sichergestellt werden, Arbeitsgeschwindigkeit sowie die verfügbaren Programme müssen also identisch sein.²⁰ Vor diesem Hintergrund erscheint die Durchführung elektronischer Prüfungen unter Verwendung der eigenen Computer der Prüflinge sehr problematisch,²¹ zumal bei der Nutzung des eigenen Computers auch Manipulationsversuche schwieriger bzw. kaum zu unterbinden sein dürften. Letzteres ist – wiederum im Hinblick auf die das gesamte Prüfungsverfahrenrecht dominierende Chancengleichheit – jedoch unbedingt zu gewährleisten.

Die zur *Vermeidung von Täuschungen* und Manipulationen notwendige Prüfungsaufsicht sollte auf herkömmliche Weise durch *Aufsichtspersonal* erfolgen. Zwar ist auch denkbar, lediglich Videoaufzeichnungen anzufertigen; allerdings dürfte ein solches Anfertigen von Videoaufzeichnungen einen zusätzlichen Stressfaktor für die Prüflinge darstellen, so dass Zweifel an der Verhältnismäßigkeit derartiger Maßnahmen im Vergleich zu einer Aufsicht durch Aufsichtspersonal gegeben sind. Zudem würden sich bei einer Videoaufzeichnung auch datenschutzrechtliche Probleme stellen.²²

Bereits vor Beginn der Prüfung ist – ähnlich wie bei der Prüfungsanmeldung – sicherzustellen, dass die Prüfungsleistung eindeutig einem Prüfling zuzuordnen ist;

¹⁸ Zum Ganzen Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (24 f.).

¹⁹ Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (27); Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (316).

²⁰ Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 37 f.

²¹ Hierzu http://ep.elan-ev.de/wiki/Rechtsfragen#cite_ref-16, Stichwort: „Chancengleichheit“.

²² Hierzu http://ep.elan-ev.de/wiki/Rechtsfragen#cite_ref-16, Stichwort: „Aufsicht.“

dies sollte u. a. durch eine Überprüfung der *Identität des Prüflings* bei Antritt zur Prüfung etwa durch Vorlage eines Ausweispapieres bei der Aufsicht erfolgen.²³

Während der Durchführung einer elektronischen Prüfungsarbeit sollte so oft wie möglich der aktuelle Stand der Prüfungsarbeit *abgespeichert* werden, um bei Systemstörungen einen Verlust der maßgeblichen Prüfungsleistungen zu vermeiden.²⁴ Auch sollte der *Verlauf* der Prüfung *dokumentiert* werden, um Rügen des Prüflings entgegentreten zu können, dass das elektronische Prüfungssystem bereits gefertigte Prüfungsleistungen nicht abgespeichert oder fehlerhaft gelöscht habe.²⁵ Sollten Systemstörungen während der Prüfung eintreten, können Ausgleichsmaßnahmen etwa in Form einer Schreibzeitverlängerung ausreichend sein; sollten derartige Ausgleichsmaßnahmen jedoch nicht in Betracht kommen, ist in letzter Konsequenz die Prüfung abzubrechen und zu wiederholen.²⁶

Personenbedingte Beeinträchtigungen sind, sofern ein derartiger *Nachteilsausgleich* nicht zu einer prüfungsrechtlich unzulässigen Überkompensation führt, durch geeignete Maßnahmen auszugleichen; so kommen zum Beispiel Bearbeitungszeitverlängerungen bei Schreibhinderungen in Frage, im Einzelfall können auch besondere prüflingsspezifische Konfigurationen des Computers (etwa eine größere Schrift bei Sehschwäche) als Ausgleichsmaßnahmen in Betracht kommen.²⁷

Nach der Prüfung

Ist die Prüfungsleistung erbracht, muss diese bewertet werden. Insbesondere bei elektronischen Prüfungen stellt sich die Frage, ob das Ergebnis der Prüfung auch automatisiert ermittelt werden darf. Im Hinblick auf das etwa in § 6a des Bundesdatenschutzgesetzes – BDSG – normierte Verbot der automatisierten Einzelfallentscheidung begegnet eine solche *automatisierte Korrektur* jedoch datenschutzrechtlichen Bedenken.²⁸ Allerdings kommt eine Rückausnahme von diesem Verbot etwa dann in Betracht, wenn die Wahrung der berechtigten Interessen des Betroffenen durch geeignete Maßnahmen gewährleistet ist und die verantwortliche Stelle dem

²³ Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (25); Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 36.

²⁴ Dazu Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (26); Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 17; s. auch Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (317).

²⁵ Vgl. hierzu Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (319).

²⁶ Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (27); s. auch Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (317 f.).

²⁷ Hierzu detailliert Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 258 ff.

²⁸ Dazu ausführlich Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (22 f.).

Betroffenen die Tatsache des Vorliegens einer automatisierten Entscheidung mitteilt sowie auf Verlangen die wesentlichen Gründe dieser Entscheidung benennt und erläutert (so § 6a Abs. 2 Nr. 2 BDSG). Hier könnte eine derartige „geeignete Maßnahme“ im datenschutzrechtlichen Sinne darin zu erblicken sein, dass im Prüfungsrecht ein aus Art. 12 Abs. 1 GG abgeleiteter, also verfassungsunmittelbarer Anspruch auf Durchführung eines Überdenkungsverfahrens²⁹ besteht.³⁰

Abgesehen von dieser datenschutzrechtlichen Problematik ist jedoch aus Sicht des Prüfungsrechts sicherzustellen, dass der Prüfer die Prüfungsleistung selbst, unmittelbar und vollständig zur Kenntnis nehmen muss und aus eigener Sicht selbständig beurteilt.³¹ Dieses Erfordernis schließt bei geschlossenen Aufgabentypen (also etwa bei Multiple-Choice-Aufgaben³²) eine automatisierte Ermittlung des Prüfungsergebnisses nicht aus, weil hier die eigentliche Prüfertätigkeit vorverlagert ist und in der Auswahl des Prüfungsstoffs, der Stellung der Aufgaben und der Konkretisierung der richtigen und falschen Antworten besteht. Die (gleichsam mechanische) Auswertung mag dann durch einen Computer erfolgen.³³ Unvereinbar mit dem Grundsatz, dass der Prüfer die *Prüfungsleistung selbst und eigenständig bewerten* muss, ist hingegen der Einsatz von (teil-)automatisierten Bewertungen bei offenen Aufgabenstellungen.³⁴ Hier kann zwar bei Kurztext-Antworten ggf. auf entsprechende Computer-Auswertungsprogramme³⁵ zurückgegriffen werden, deren Ergebnisse dürfen jedoch nur als gänzlich unverbindliche Hilfestellung bei der ohne Ausnahme zu fordernden persönlichen und eigenständigen Bewertung durch den Prüfer herangezogen werden.

Schließlich ist sicherzustellen, dass die Prüfungsleistung mindestens bis zur Bestandskraft der jeweiligen Prüfungsentscheidung *archiviert* wird und eine *Einsicht* durch den Prüfling möglich ist. Dies kann durch einen Ausdruck der elektronischen Prüfungsdateien oder in elektronischer Form geschehen, wobei bei der elektronischen Speicherung Sicherungsmaßnahmen gegen eine nachträgliche Veränderung

²⁹ Hierzu detailliert Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 783 ff.

³⁰ So Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (23); http://ep.elan-ev.de/wiki/Rechtsfragen#cite_ref-16, Stichwort: „Automatisierte Korrektur.“

³¹ Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 320 m. w. N.

³² Zu den im Zusammenhang mit Multiple-Choice-Aufgaben zu berücksichtigenden Besonderheiten s. detailliert Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 588 ff.

³³ Niehues / Fischer / Jeremias, Prüfungsrecht, 6. Aufl. 2014, Rdn. 589.

³⁴ S. auch Zimmermann, Online-Prüfungen – Rechtliche Grundlagen und Konzeptionierung, WissR 2012, 312 (322).

³⁵ S. dazu Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 22.

der Daten getroffen werden müssen und zudem datenschutzrechtliche Probleme zu bewältigen sein können.³⁶

Fazit

Bei der Durchführung von elektronischen Prüfungen ist zuvorderst sicherzustellen, dass eine ausreichende Rechtsgrundlage für diese Art der Prüfung gegeben ist. Findet sich eine solche Rechtsgrundlage etwa in der Prüfungsordnung, so folgen die prüfungsrechtlichen Anforderungen an die Durchführung einer elektronischen Prüfung weitestgehend den allgemeinen prüfungsrechtlichen Anforderungen. Hierbei ist stets die Bedeutung der Berufsfreiheit (Art. 12 Abs. 1 GG) und der Chancengleichheit (Art. 3 Abs. 1 GG) zu gewärtigen.

Vita

Der Autor ist seit 2007 Richter in Berlin und seit 2008 – unterbrochen durch eine zweijährige Abordnung als Wissenschaftlicher Mitarbeiter an das Bundesverwaltungsgericht – am Verwaltungsgericht Berlin tätig.

³⁶ Näher hierzu http://ep.elan-ev.de/wiki/Rechtsfragen#cite_ref-16, Stichwort: „Archivierung“ sowie „Akteneinsicht und Aufbewahrungspflicht“; Kalberg, Rechtsfragen computergestützter Präsenzprüfungen im Antwort-Wahl-Verfahren, DVBl. 2009, 21 (28); Vogt / Schneider, E-Klausuren an Hochschulen, 2009, S. 24 f., 37.

Kompetent und sicher: Online-Prüfungen mit Virtueller Desktop Infrastruktur und Safe Exam Browser an der ETH Zürich

Einleitung

Zahllose Arbeitsprozesse sind heute direkt von Informatikmitteln abhängig. Ingenieurswesen, Statistik, Biochemie, Gestaltung, Design oder die Informatik selbst sind ohne Computer heute undenkbar: Er hat Papier und Stift als wichtigstes Arbeitswerkzeug abgelöst. Auch die Lehre an Hochschulen ist heute weitgehend digitalisiert – eine Entwicklung, die auch vor dem Prüfungsbereich nicht Halt macht. Die Zahl von Hochschulen, welche Online-Prüfungen in der einen oder anderen Form anbieten, nimmt stetig zu. Dabei wird der wesentliche Mehrwert von Online-Prüfungen meist in einer vereinfachten Abwicklung und Korrektur von Prüfungen sowie der damit einhergehenden Zeitersparnis für die Examinatoren verortet. Der Computer dient als Hilfsmittel zur effizienteren Darstellung und Durchführung von Prüfungen, welche aber inhaltlich im Vergleich zu ihren vormaligen Papierversionen weitgehend unverändert bleiben. Statt mit Papier und Stift werden dieselben Kurzantwortaufgaben neu per Mausklick beantwortet oder dieselben Essays neu mittels Computertastaturen verfasst.

Die ETH Zürich führte ihre ersten Online-Prüfungen 2007 bis 2009 im Rahmen des Projekts „Online-Prüfungen an der ETH Zürich“ durch. Seit 2010 werden Online-Prüfungen vom Bereich Lehrentwicklung und -technologie (LET) zusammen mit den Informatikdiensten als reguläre, hochschulweite Dienstleistung angeboten (vgl. Piendl, Halbherr & Schneider, 2014). Dabei zeigte sich schon früh, dass es einen Bedarf an Online-Prüfungen gibt, welche über die oben beschriebenen Funktionalitäten im Rahmen herkömmlicher Prüfungen hinausgehen. Examinatoren wünschten sich, dass ein Weg gefunden wird, Studierenden einen prüfungssicheren Zugriff auf fachspezifische Software zu gewähren. In Matlab sollen Programmieraufgaben bearbeitet, in R-Studio Statistiken gerechnet werden, Vorlesungsskripte als PDF-Dateien verfügbar sein oder Pflanzenarten mittels einer eigens entwickelten E-Learning-Software bestimmt werden. Statt in einer Statistikprüfung praxisfern und auf Umwegen mittels Kurzantwort- oder Essay-Aufgaben zu prüfen, ob ein statistisches Verfahren verstanden wurde, soll authentisch und praxisnah geprüft werden können: Die Studierenden erhalten eine statistische Fragestellung zusam-

men mit einem entsprechenden Datensatz und sollen ihre erworbenen Kenntnisse in der direkten Anwendung unter Beweis stellen.

Damit solche Prüfungsszenarien praxistauglich und prüfungssicher sein können, müssen sie einer Reihe von Anforderungen genügen, insbesondere: Betrugssicherheit, Benutzerfreundlichkeit, Zuverlässigkeit, Skalierbarkeit sowie Flexibilität. In diesem Beitrag stellen wir eine technische Infrastruktur für Online-Prüfungen mit Drittapplikationen vor, welche auf einer virtuellen Desktopinfrastruktur, Safe Exam Browser sowie dem Learning Management System Moodle basiert. Die ETH Zürich führt seit Dezember 2012 solche Online-Prüfungen mit VDI, SEB und Moodle, kurz „VDI-Prüfungen“, durch.

Prüfungsdidaktische Überlegungen

Eine wichtige Zielsetzung der ETH Zürich für VDI-Prüfungen – wie auch für Online-Prüfungen allgemein – ist die Verbesserung der Qualität von Prüfungen (vgl. Halbherr et al., 2014). Prüfungsqualität soll dabei an zwei Hauptkriterien bemessen werden: Prüfungen sollen erstens valide sein und zweitens Anreize schaffen, welche studentisches Lernen positiv beeinflussen. Validität ist das Ausmaß der Übereinstimmung zwischen dem, was eine Prüfung zu messen beansprucht und dem, was tatsächlich gemessen wird. Zahlreiche wichtige Anliegen im Prüfungskontext, wie Gleichbehandlung, Fairness, Objektivität oder die Angemessenheit der Aufgaben sind Teilmengen der Validität (vgl. Newton & Shaw, 2014). Prüfungen haben einen sehr starken Einfluss darauf was, wie und wieviel Studierende lernen (vgl. Sambell & McDowell, 1998). Die allgegenwärtige Frage: „Kommt das in der Prüfung?“ veranschaulicht diesen Zusammenhang sehr schön. Aus Sicht der Studierenden ist die Prüfung das Curriculum und das eigene Lernen wird deshalb vor allem Anderen auf die erwartete Prüfung ausgerichtet. Eine qualitativ hochstehende Prüfung soll sich deshalb auch durch positive Auswirkungen auf das Lernverhalten Studierender auszeichnen.

Zur Gewährleistung der Prüfungsqualität sollen drei didaktische Konzepte hervorgehoben werden: Constructive Alignment, Kompetenzorientierung sowie authentisches Prüfen. *Constructive Alignment* (Biggs, 1996) postuliert das Folgende. Erstens ist Lernen als aktiver Prozess des Lernenden zu begreifen, bei welchem neues Wissen und Verständnis geschaffen wird, indem neue Erfahrungen mit bestehenden Kenntnissen in Beziehung gebracht werden (*constructive*). Zweitens setzt erfolgreiche Lehre eine Übereinstimmung von Lernzielen, Lehr- und Lernaktivitäten sowie Leistungskontrollen voraus (*alignment*). *Constructive Alignment* soll als Leitprinzip die Gestaltung qualitativ hochwertiger Kurse erleichtern. Die *Kompetenzorientierung* ist dem Alignmentgedanken nahe verwandt. Weinert (2001, S. 27-28) definiert Kompetenzen als „die bei Individuen verfügbaren oder durch sie erlernbaren *kogni-*

tiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ Die Kompetenzorientierung grenzt sich dabei vor allem gegenüber einem früheren, behavioristisch geprägten Verständnis von Lernen ab. Dieses verstand Lernen allein als das Vermitteln von Wissen. Bei einem kompetenzorientierten Ansatz hingegen wird Lernen als der Erwerb von Kenntnissen – Wissen miteingeschlossen – verstanden, welche situationsbezogen zur Lösung bestimmter Probleme befähigen. *Authentisches Prüfen* (vgl. Gielen, Dochy & Dierick, 2003) schließlich ist das aus den obigen Überlegungen abgeleitete Postulat, dass Prüfungsqualität im Sinne der Validität und Lerndienlichkeit gewährleistet werden kann, indem Prüfungsaufgaben „authentisch“ gestaltet werden. Die Authentizität ist dabei als der Grad der Übereinstimmung von Prüfungsaufgaben mit „realen“ Problemstellungen gemäß Lernzielen zu verstehen. Solche „realen“ Problemstellungen können sich dabei sowohl auf Zielkompetenzen in einer späteren beruflichen oder akademischen Tätigkeit beziehen als auch auf notwendige Zwischenschritte im Kompetenzerwerb innerhalb des aktuellen Bildungskontexts.



Abbildung 1: Eine Prüfungsgestaltung gemäß den Zielsetzungen von *Constructive Alignment*, Kompetenzorientierung und *authentischem Prüfen* ist aufgrund der in den meisten Fachgebieten weit vorangeschrittenen Digitalisierung mit Papier und Stift häufig nicht mehr möglich.

Aufgrund der in vielen Fachgebieten weit vorangeschrittenen Digitalisierung muss angesichts dieser Überlegungen ein weiteres Prüfen mit Papier und Stift kritisch hinterfragt werden (vgl. Abbildung 1). Ein authentisches Prüfen ist hier kaum mehr möglich und so entfernen sich sowohl die in Prüfungen de facto erhobenen Kompetenzen als auch die zur Prüfungsvorbereitung von Studenten eingeübten Kompetenzen zusehends von dem, was eigentlich gemäß Lernzielen gelernt werden soll. Es ist deshalb erforderlich, dass Studierende auch in Prüfungen mit fachspezifischer Software sowie weiteren relevanten digitalen Werkzeugen arbeiten können.

Technische Anforderungen

Damit sich kompetenzorientierte Online-Prüfungen mit Drittapplikationen im Prüfungsalltag bewähren können, muss eine technische Umsetzung gefunden werden, welche den folgenden Anforderungen genügt: Betrugssicherheit, Benutzerfreundlichkeit, Zuverlässigkeit, Skalierbarkeit und Flexibilität.

In der Regel ist ein Zugriff Studierender auf beliebige Software und Ressourcen in einer Online-Prüfung nicht erwünscht. Aber selbst in einer Open-Book-Prüfung, bei welcher ein Zugriff auf beliebige Software, Dateien sowie alle über das Internet verfügbaren Ressourcen explizit erlaubt wäre, muss gewährleistet werden, dass die von Kandidaten abgegebene Arbeit eine Eigenleistung darstellt und nicht einfach an eine Drittperson delegiert wurde. Es muss also sichergestellt werden, dass Studierende während der Prüfung lediglich explizit erlaubte Ressourcen verwenden können. Diese Aufgabe an die Prüfungsaufsicht allein zu delegieren ist kaum praktikabel, weshalb eine zusätzlich technische Absicherung benötigt wird. Diese soll den Zugriff auf erlaubte Ressourcen ermöglichen und gleichzeitig auf nicht erlaubte verhindern.

Die Prüfungsleistung Studierender soll wiedergeben, wie gut Lernziele erreicht wurden. Stressresistenz oder die Fähigkeit sich in einer unvertrauten Arbeitsumgebung zurechtfinden zu können, sollen hingegen möglichst keinen Einfluss auf die Prüfungsleistung haben. Es ist deshalb darauf zu achten, dass abgesicherte Prüfungsumgebungen ein Arbeiten ermöglichen, welches so wenig wie möglich von der Arbeit auf einem unabgesicherten Rechner abweicht. Wo die abgesicherte Prüfungsumgebung von einem unabgesicherten Rechner dennoch abweicht, soll deren Handhabung so einfach und intuitiv wie möglich sein. Zusätzlich sollen die Kandidaten die Gelegenheit erhalten, sich vor der Prüfung mit den abgesicherten Umgebungen vertraut zu machen.

An der ETH Zürich werden unterschiedlichste Kombinationen von Applikationen und Konfigurationen im Prüfungskontext verwendet. Das System muss daher flexibel genug sein, um diese Szenarien effizient abdecken zu können. Da im Prüfungsbetrieb Prüfungen mit unterschiedlichen Konfigurationen gleichzeitig oder direkt aufeinanderfolgend stattfinden können, muss einfach und zuverlässig zwischen verschiedenen Konfigurationen gewechselt werden können. Aufgrund der Tatsache, dass bei Prüfungen für die Kandidaten viel auf dem Spiel steht, müssen alle beteiligten technischen Systeme bezüglich Skalierung, vor allem aber in Bezug auf Zuverlässigkeit, besonders hohen Ansprüchen genügen. Kommt es trotz aller Vorkehrungen zu einer technischen Störung bei einer Prüfung, soll die Nachvollziehbarkeit der Ereignisse gewährleistet sein.

Technische Umsetzung

Die technische Infrastruktur für kompetenzorientierte Online-Prüfungen mit Drittapplikationen an der ETH Zürich basiert auf drei Hauptkomponenten: Einer virtuellen Desktopinfrastruktur (VDI), Safe Exam Browser (SEB) sowie dem Learning Management System (LMS) Moodle. Eine virtuelle Desktopinfrastruktur, teilweise auch Desktopvirtualisierung genannt, ist eine technische Möglichkeit, ein für den Endanwender gedachtes Betriebssystem, wie zum Beispiel Microsoft Windows, in einem Rechenzentrum zu virtualisieren, wie dies für Serverbetriebssysteme seit vielen Jahren erfolgreich durchgeführt wird. VDI bietet in Unternehmen unter anderem mögliche Verbesserungen bei der Clientverfügbarkeit, IT-Sicherheit sowie bei IT-Kosten. Der Safe Exam Browser ermöglicht eine gegen Betrug abgesicherte Durchführung von Online-Prüfungen mit Hilfe von Quiz-Modulen von Learning Management Systemen. Hierzu blockiert eine Komponente der Applikation sämtliche nicht erlaubten Möglichkeiten des Computers, während eine andere Komponente die Darstellung der Prüfung des Learning Management Systems übernimmt.

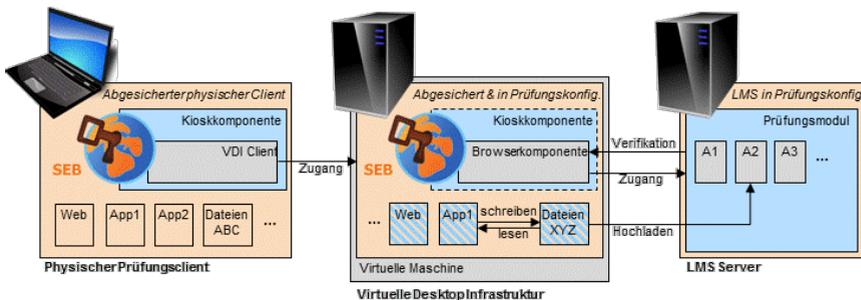


Abbildung 2: Grundprinzip der Absicherung einer VDI-Prüfung

Abbildung 2 stellt das Grundprinzip der Absicherung einer Online-Prüfung mit Hilfe einer VDI-Umgebung und SEB dar. Der physische Prüfungsclient wird mit SEB abgesichert. SEB verhindert dabei mit seiner Kiosk-Komponente den Zugriff auf sämtliche auf dem Rechner installierten Applikationen und gespeicherten Dateien. SEB startet einen VDI Client, mit welchem auf eine virtuelle Maschine zugegriffen wird. Auf der virtuellen Maschine ermöglichen diverse abgesicherte Applikationen die Durchführung einer Prüfung. Der Safe Exam Browser auf der virtuellen Maschine erfüllt dabei drei Funktionen: Er sichert die virtuelle Maschine ab, dient als Webbrowser um auf ein LMS zugreifen zu können und er gewährt den Zugang zu den benötigten Applikationen. Die Applikationen wiederum können auf Dateien zugreifen, diese editieren sowie abspeichern. Im Prüfungsmodul des LMS werden die Prüfungsaufgaben dargestellt, erstellte Dateien durch Hochladen abgegeben

und es stehen klassische Aufgabenformate wie Multiple Choice oder Essay zur Verfügung.

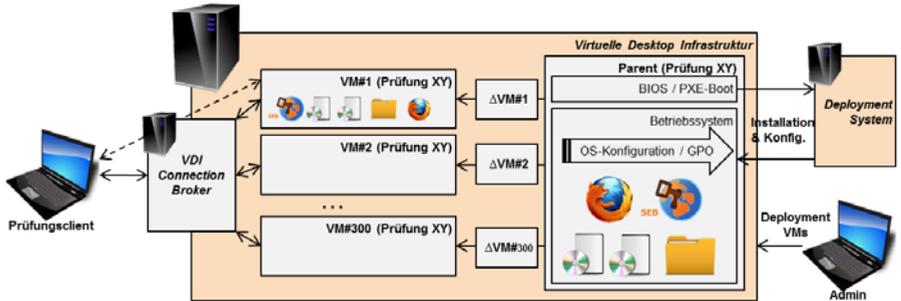


Abbildung 3: Schematische Darstellung der Software-Komponenten einer VDI-Prüfungsumgebung

Abbildung 3 zeigt auf, wie die unterschiedlichen Softwarekomponenten der Infrastruktur für VDI-Prüfungen miteinander verknüpft sind. Mittels des Prüfungsclients wird auf eine virtuelle Maschine zugegriffen, die eine Kopie des entsprechenden Parents ist. Zunächst wird auf jedem Parent mittels des Softwareverteilungssystems, auch „Deployment-System“, Baramundi das Betriebssystem installiert. Hierzu wird ein Netzwerkboot / PXE via BIOS durchgeführt. Ebenfalls werden via Baramundi die Applikationsinstallationen (z. B. Matlab, R-Studio) vorgenommen. Mittels Group Policy Objects (GPO) werden die Sicherheitsanforderungen umgesetzt. Sobald ein Parent fertig konfiguriert ist, wird der Stand in einem Snapshot abgespeichert. Dieser Snapshot wird dann auf die virtuellen Maschinen als Deltaimage ($\Delta VM\#n$) verteilt. Die ETH Zürich setzt für die Desktopvirtualisierung gegenwärtig auf die Produktreihe *Horizon View* der Firma VMware.

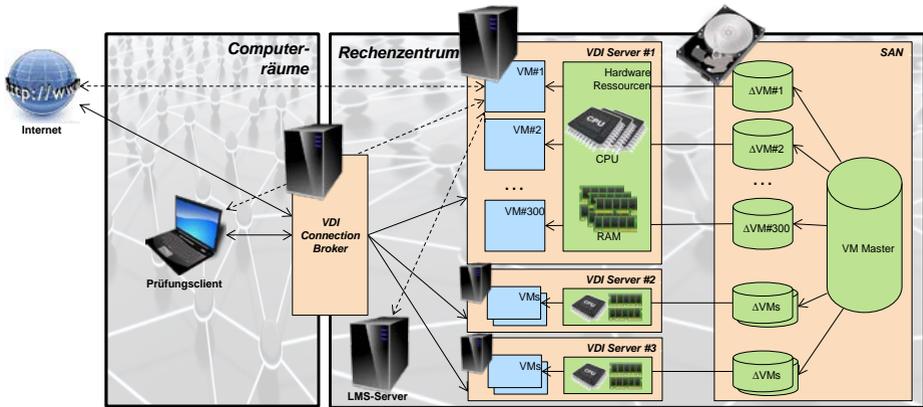


Abbildung 4: Schematische Darstellung der verwendeten Systeme der VDI-Prüfungsumgebung

Die in Abbildung 4 schematisch dargestellte Hardware ist primär in zwei Bereiche aufgeteilt. Auf der linken Seite der Abbildung findet sich der Prüfungsclient im Computerraum, ein physisch vorhandener Computer. Auf der rechten Seite ist die Serverinfrastruktur im Rechenzentrum dargestellt. Auf den physischen Servern VDI Server #1, VDI Server #2 und VDI Server #3 werden gegenwärtig bis zu 300 virtuelle Maschinen hochverfügbar betrieben. Die Daten der virtuellen Maschinen werden als Delta zu einem Master-Image auf einem Hochleistungs-SAN (Storage Area Network) gespeichert. Die Verbindung zwischen dem Prüfungsclient und einer virtuellen Maschine wird über den VDI Connection Broker hergestellt. Auf dem Prüfungsclient wird dazu die Applikation VMware Horizon View Client benötigt. Die Verbindung zwischen virtueller Maschine und LMS-Servern bzw. Internet erfolgt nicht über den VDI Connection Broker, sondern kann direkt hergestellt werden.

Tabelle 1: Hardware-Ausstattung pro ESXi-Server

Prozessor	4 x Intel E5-4640
Anzahl Prozessorkerne (physikalisch)	32
SMT aktiviert	Ja
RAM-Größe	768 GB
Netzwerkkarten	4 x 10Gbit Ethernet

Die Serverinfrastruktur besteht aus drei ESXi-Servern. Tabelle 1 gibt eine Übersicht der Ausstattung dieser Server. Sämtliche Server laufen aktuell mit VMware ESXi, Version 6.0. Die Server befinden sich in einem Rechenzentrum der ETH Zürich und werden von den zentralen Informatikdiensten der Hochschule betreut.

Die ETH Zürich hatte zu Beginn des VDI-Projektes die folgende Infrastruktur in Betrieb: Einen großen Computerraum mit 170 Prüfungsarbeitsplätzen, vier mittelgroße Computerräume mit insgesamt 136 Prüfungsarbeitsplätzen sowie drei kleine Computerräume mit weiteren 53 Prüfungsarbeitsplätzen. Jeder Arbeitsplatz ist mit einem zentral verwalteten PC sowie dazugehöriger Peripherie (Maus, Tastatur, Bildschirm) ausgestattet. Jeder Computer ist mittels Gigabit-Ethernet mit dem Netzwerk der ETH Zürich verbunden. Weitere zentrale Netzwerkkomponenten wie beispielsweise DNS, DHCP, AD / LDAP, Router, Switches etc. waren ebenfalls bereits vorhanden. Das VDI-Projekt war zwingend auf diese bestehende Infrastruktur angewiesen.

Um die Zuverlässigkeit einer VDI-Prüfung sicherstellen zu können, setzt die ETH Zürich verschiedene technische und organisatorische Möglichkeiten ein. Beispielsweise werden an technischen Schlüsselstellen Redundanzen eingesetzt. Ebenfalls investiert die ETH Zürich viele Ressourcen in das Testen der Systeme sowie das Training des technischen Personals. VDI-Prüfungen an der ETH Zürich können gegenwärtig von bis zu 274 Kandidaten gleichzeitig abgelegt werden, finden aber auch in kleineren Veranstaltungen mit 20 oder weniger Kandidaten statt. Das VDI-System muss also sowohl mit sehr wenig Aufwand einsatzbereit sein als auch eine große Anzahl gleichzeitiger Zugriffe bewältigen können. Die Leistungsfähigkeit der Systeme wird mit Lasttests regelmäßig überprüft.

Betrieb

Zur Vorbereitung einer Prüfung muss der Parent installiert, konfiguriert und getestet werden. Danach kann die benötigte Anzahl virtueller Maschinen verteilt werden. Der Arbeitsaufwand für die Verteilung liegt im Bereich einer Stunde. Bevor die VDI-Umgebung zur Verfügung stand, wurde ein Versuch gestartet, ähnliche Möglichkeiten auch auf physikalischen Maschinen anzubieten. Dabei wurden diverse Schwierigkeiten und Einschränkungen festgestellt, welche unter anderem zum Entscheid zur Durchführung des VDI-Projektes beigetragen hatten. Die Schwierigkeiten waren wie folgt:

Zeit und Aufwand: Die Computerräume werden vor und nach einer Prüfung von Studierenden genutzt, weshalb die Zeit zum Starten der Prüfungskonfiguration maximal eine Stunde in Anspruch nehmen darf. Nach der Prüfung muss ein Raum ebenfalls innerhalb einer Stunde wieder freigegeben werden können. Mit der VDI-Prüfungsumgebung halten wir diese Zeiträume problemlos ein, während die Umkonfiguration von physikalischen Maschinen mehrere Stunden benötigte und sich zudem als störungsanfällig erwies.

Kongruenz der Systeme: Es gibt keine skalierbare Möglichkeit festzustellen, ob sich sämtliche physikalischen Computer im gleichen Zustand befinden und exakt die

gleiche Software mit der exakt gleichen Konfiguration verwenden. In der Praxis zeigte sich, dass der Zustand der physikalischen Maschinen im Allgemeinen stark voneinander abwich. Diese Heterogenität führte zu zahlreichen verschiedenen, häufig auftretenden sowie schwer nachvollziehbaren Zwischenfällen im Betrieb der physikalischen Maschinen. Da die virtuellen Maschinen vor jeder Prüfung aus dem entsprechenden Parent neu erstellt werden, ist sichergestellt, dass ihr Zustand nicht nur in der Theorie, sondern auch tatsächlich identisch ist.

Nebenwirkungen aufgrund von Sicherheitsbeschränkungen: Um die Sicherheit einer Prüfung zu gewährleisten, müssen die Netzwerkverbindungen eingeschränkt werden, zum Beispiel indem die DNS-Auflösung blockiert wird oder ein Proxy den Netzwerkverkehr filtert. Diese Maßnahmen können Nebenwirkungen auf den restlichen Betrieb der ETH Zürich haben, weshalb sie nicht eingesetzt werden dürfen. Gleichzeitig werden diese aber für die Absicherung benötigt. Da die virtuellen Maschinen im Betrieb nicht wie die physischen Prüfungscomputer zwischen Prüfungs- und Normalkonfiguration hin- und herwechseln müssen, kann diese Absicherung durch eine entsprechende Konfiguration der virtuellen Maschinen gewährleistet werden.

Sind die virtuellen Maschinen verteilt, laufen VDI-Prüfungen im Wesentlichen gleich ab wie „normale“ browserbasierte Online-Prüfungen (vgl. Halbherr et al. 2014). Eine Stunde vor Prüfungsbeginn werden die Computer automatisiert neu gestartet. Nach dem Neustart wird die Konfiguration der Rechner angepasst, spezielle Prüfungsaccounts werden eingeloggt, SEB sowie der View Client werden gestartet und dieser loggt sich mit dem verwendeten Prüfungsaccount wiederum in die virtuelle Maschine ein. Dieser gesamte Ablauf findet automatisiert statt und kann ganzen Computerräumen oder einzelnen Maschinen zugewiesen werden. Die Studierenden finden die Maschinen im Prüfungszustand vor. Auf dem Bildschirm sehen sie eine Portalseite mit Links zu den aktuellen Online-Prüfungen im Prüfungsmodul des LMS. Die Portalseite wird dabei von der Browserkomponente des SEB auf der virtuellen Maschine dargestellt. Im Unterschied zu reinen browserbasierten Online-Prüfungen wird den Studierenden vor Beginn der Prüfung Zeit gewährt, um sich mit den abgesicherten virtuellen Maschinen sowie ihrer Handhabung vertraut machen zu können. Etwaige Fragen werden beantwortet und Unklarheiten geklärt. In der Praxis hat sich gezeigt, dass hierfür fünf bis maximal zehn Minuten ausreichend Zeit sind. In der Prüfung tauchen anschließend kaum mehr Fragen zur Handhabung auf. Es wird deshalb inzwischen auf die Durchführung von Probeprüfungen verzichtet.

Jede VDI-Prüfung wird von technischem vor-Ort-Support begleitet. Der vor-Ort-Support koordiniert Maßnahmen im Falle technischer Pannen, wie zum Beispiel das Umsetzen von Kandidaten bei dem Ausfall eines Prüfungsclients. Im Hintergrund steht der Second-Level-Support für die involvierten Infrastrukturen auf Abruf

und überwacht die entsprechenden Systeme. Die folgenden Objekte der Serverinfrastruktur werden während einer Online-Prüfung automatisiert überwacht: CPU-Auslastung, RAM-Auslastung, Disk-IO, Zertifikate, Netzwerkverfügbarkeit, DNS-Auflösung. Die Überwachung erfolgt durch die beiden Monitoringsysteme Nagios und GroundWork. Neben den automatischen Überwachungen, werden auch die Logdateien des Serversystems ständig überprüft. Seit Sommer 2014 setzt die ETH Zürich zur Kommunikation zwischen vor-Ort-Support und Second-Level-Support während Prüfungen auf Funkgeräte. Online-Prüfungen an der ETH Zürich finden häufig auf mehrere Computerarbeitsräume verteilt statt, welche wiederum über bis zu vier verschiedene Stockwerke verteilt sind. Während eines Fehlerfalles müssen die für die Online-Prüfung technisch Verantwortlichen schnell und effizient kommunizieren können.

Diskussion & Ausblick

Die ETH Zürich setzt das VDI-System seit Dezember 2012 ein. Bisher (Stand April 2015) wurden 28 Prüfungen mit insgesamt 1'300 Prüfungsleistungen durchgeführt. Im Herbstsemester 2014 fanden rund 10% aller am Computer abgelegten Prüfungen auf der VDI-Umgebung statt. Die bisherigen Erfahrungen mit dem System sind äußerst positiv. Durch den generischen Ansatz können viele Anforderungen der Dozierenden einfach und effizient umgesetzt werden. Die größte aufgetretene Schwierigkeit bestand darin, die abgesicherten Arbeitsumgebungen für die Studierenden ausreichend einfach bedienbar und benutzerfreundlich zu gestalten. So bekundet beispielsweise noch immer ein zwar kleiner, dafür aber erstaunlich stabiler Prozentsatz der Prüfungskandidaten Mühe mit dem fehlerfreien Hochladen von Dateien in unser Learning Management System Moodle. Aufgrund der relativ neuen Technologie sowie der hohen Komplexität des Gesamtsystems sind ursprünglich erwartete technische Schwierigkeiten bisher ausgeblieben. Das System läuft sehr stabil und hat im gesamten Betrieb kaum technischen Support benötigt.

Die ETH Zürich möchte die VDI-Umgebung für die Benutzung mit CAD-Applikationen bereit machen. Dies bedingt allerdings eine deutliche Steigerung der möglichen Grafikleistung. Ein Projekt, das im Herbst 2014 gestartet ist, versucht dies mittels der Verwendung von Nvidia Grid K1 und K2 Grafikkarten zu ermöglichen. Diese Technologie steckt gegenwärtig noch im Anfangsstadium und ist zum jetzigen Zeitpunkt sehr kostenintensiv. Ein weiteres Hochskalieren der Infrastruktur für VDI-Prüfungen wird hingegen erst erforderlich sein, wenn die Kapazität der Computerprüfungsräume weiter erhöht wird. Schließlich und vor allem soll der Service VDI-Prüfungen qualitativ bessere Prüfungen ermöglichen. Der Bereich Lehrentwicklung und -technologie steht diesbezüglich in engem Kontakt und Austausch mit Dozierenden und Departementen (Fakultäten) um die weitere Entwicklung der VDI-Prüfungspraxis didaktisch unterstützen und mitgestalten zu können.

Referenzen

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364.

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 (Hrsg.), *Optimising new modes of assessment: In search of qualities and standards* (S. 37-54). Dordrecht: Kluwer Academic Publishers.

Halbherr, T., Reuter, K., Schneider, D., Schlienger, C. & Piendl, T. (2014). Making examinations more valid, meaningful and motivating: The Online Exams Service at ETH Zurich. *EUNIS Journal of Higher Education IT*, 2014-1.

Newton, P. & Shaw, S. (2014). *Validity in educational and psychological assessment*. Los Angeles: SAGE.

Piendl, T., Halbherr, T. & Schneider, D. (2014). Online-Prüfungen an der ETH Zürich: Vom Projekt zum Service. In T. Škerlak, H. Kaufmann & G. Bachmann (Hrsg.), *Lernumgebungen an der Hochschule: Auf dem Weg zum Campus von morgen*. Münster: Waxmann.

Sambell, K. & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, 23(4), 391-402.

Weinert, F. E. (2002). *Leistungsmessungen in der Schule*. Weinheim: Beltz.

Vita

Tobias Halbherr ist Fachexperte für Prüfungen im Stabsbereich Lehrentwicklung und -technologie (LET) der ETH Zürich und leitet den Service „Online-Prüfungen.“ Er hat an der Universität Zürich und der ETH Zürich Psychologie und Informatik studiert. Frühere Tätigkeiten beinhalten Forschung zu Human Factors in der Flugsicherheit, Objekterkennung und computergestütztem Lernen in der Visual Cognition Research Group (VICOREG) der Universität Zürich sowie im Center for Adaptive Security Research (CASRA); Softwareentwicklung und quantitative Analysen im Engineering Treasury der Zürcher Kantonalbank; selbständige Tätigkeit in Design und Betrieb von live-Video Installationen.

Kai Reuter ist Systemadministrator im Stabsbereich Lehrentwicklung und -technologie (LET) der ETH Zürich. Neben dem Betrieb der Learning Management Systeme ist er zuständig für den Betrieb und die Weiterentwicklung der VDI-Prüfungsumgebung. Nach seiner Ausbildung zum Applikationsentwickler EFZ hat er ein Studium zum dipl. Wirtschaftsinformatiker HF abgeschlossen.

Verwendung der Bürgerkarte (Digitale Signatur) für E-Assessment: Realisierung von sicherheitstechnischen und rechtlichen Anforderungen an elektronische Prüfungen

Zusammenfassung

Im Zuge der Einrichtung eines Prüfungsraumes für elektronische Prüfungen an der Johannes Kepler Universität Linz (Österreich) wurden neben der räumlichen und technischen Ausstattung auch die organisatorischen, prüfungsadministrativen und sicherheitstechnischen Prozesse neu konzipiert. Um die sicherheitstechnischen und rechtlichen Anforderungen zu realisieren wurde eine PKI-Umgebung (Public Key Infrastructure) in einem nur für Prüfungszwecke verwendeten PC-Labor installiert. Ziel der technischen und organisatorischen Lösungen ist es, Lehrende, Studierende sowie Prüfungsaufsicht und -administration gleichermaßen zu unterstützen.

Einleitung

Mit der Einführung des Bologna-Prozesses und der damit einhergehenden Reformierung der Studienstrukturen wurde auch in Österreich auf das konsekutive dreigliedrige Studiensystem Bachelor-Master-Doktorat umgestellt. Damit wurden in den Bachelor-Studien die Studienpläne dahingehend geändert, dass es zu einer Aufspaltung in kleinere Facheinheiten mit vielen Lehrveranstaltungen und den entsprechenden Prüfungen kommt, was zu einer Bedeutungsaufwertung der Prüfungen führte (Müller, 2011). Die Messung von Kompetenzen in Form von Prüfungen ist die Voraussetzung dafür, Bildungsabschlüsse mit Blick auf die erworbenen Kompetenzen vergleichen zu können und damit das Bildungssystem durchlässiger zu machen (Reinmann, 2011). Dies dient als Ausgangspunkt für das Projekt MUSSS (Multimedia Studien Support Sozial- und Wirtschaftswissenschaften). Einerseits bietet E-Learning berufstätigen Studierenden und Studierenden mit Betreuungsaufgaben die Chance auf ein universitäres Studium und damit ein Lifelong Learning. Andererseits sollen durch dieses Programm auch Personen angesprochen werden, die örtlich disloziert von der Universität wohnen.

Im ersten Schritt des Projektes MUSSS wurden jene Lehrveranstaltungen, die von einer großen Anzahl von Studierenden des Bachelor-Studiums Wirtschaftswissenschaften benötigt werden, meist parallel zur Präsenzlehrveranstaltung als Blended-Learning-Angebot konzipiert. Je nach Lehrveranstaltungsformat und Fachinhalt werden die Lehrinhalte mit unterschiedlichen Medien (häufig kommentierte Folien, Wikis usw.) in der Lernplattform Moodle zur Verfügung gestellt. Die Präsenzphasen werden reduziert bzw. durch Webvideokonferenzen ersetzt und dort eingesetzt, wo die Interaktion der Lernenden untereinander bzw. zwischen Lehrenden und Lernenden im Mittelpunkt steht, wie es beispielsweise auch das Konzept des *Flipped Classroom*-Modells vorsieht (vgl. Bergmann & Sams, 2012; Treack, Himpsl-Gutermann & Robes, 2013). Im Sprachunterricht, wo die kommunikativen Fähigkeiten trainiert werden, wird ein höherer Präsenzanteil angeboten (bis zu einem Drittel der traditionellen Lehrveranstaltung). In Fächern, in denen deklaratives Wissen (Achtenhagen & Beathge, 2007) im Vordergrund steht und die Lernmaterialien entsprechend aufbereitet sind, gibt es keine verpflichtenden Präsenzphasen.

Am Beginn der Entwicklung von MUSSS stand vorwiegend die Neukonzeption bzw. Überarbeitung von Lehrveranstaltungen. Dabei wurden die Lehrinhalte mit digitalen Lernmedien aufbereitet. Im weiteren Ausbau von MUSSS rückte die Kompetenzüberprüfung in den Fokus. Die Überprüfung der Wissensdimension und die damit verbundene Beurteilung der Studierenden stellt bei Lehrveranstaltungen mit hohen Studierendenzahlen eine besondere Herausforderung für die Lehrenden dar. Allein schon aus der Sicht der Arbeitskapazität der Lehrenden rückt eine computerunterstützte Prüfung (E-Assessment) in den Mittelpunkt des Interesses.

E-Assessment

Mit Hilfe von Assessment-Aktivitäten können an verschiedenen Punkten des Lernprozesses Informationen über die Kenntnisse und Fähigkeiten der Lernenden gesammelt werden. Als kleinster gemeinsamer Nenner gilt die Unterscheidung zwischen *assessment for learning* vs. *assessment of learning*. Diagnostisches und formatives Assessment wird am Beginn bzw. während des Lernprozesses eingesetzt, um in den Lernprozess noch steuernd eingreifen zu können. Summative Assessments stehen am Ende des Lernprozesses und dienen der Beurteilung und Benotung der Lernenden. Sie schließen einen Lernprozess ab und haben oft selektiven Charakter. Sie können damit entscheidend für den weiteren Bildungsweg sein. Das summative Assessment steht als Teilprüfung bzw. Endprüfung am Ende einer Lehrveranstaltung. Es wird überprüft, ob der oder die Studierende das Lernziel der Veranstaltung erreicht hat und die Lehrperson bescheinigt die erbrachte Leistung. Das Hauptaugenmerk liegt dabei auf der Überprüfung eines bestimmten Kompetenzniveaus.

Unter E-Assessment versteht man eine Lernfortschrittskontrolle, die mit digitalen Medien vorbereitet, durchgeführt und nachbereitet wird. Die (teil)automatisierte Korrektur der Prüfung spielt dabei eine zentrale Rolle (Eilers, Gruttmann & Kuchen 2008). Einhergehend mit der Verbreitung von E-Learning-Szenarien werden auch Modelle des E-Assessment erprobt, die auf diese Entwicklung abgestimmt sind. Ein Beispiel hierfür ist das integrierte Modell für E-Assessment von Wesiak und Kollegen (2013), das ausgehend von den Lernzielen die unterschiedlichen Lernressourcen berücksichtigt, um unterschiedliche Formen des Assessments zu etablieren und in die Evaluation miteinzubeziehen.

Die klassische Testtheorie legt als Gütekriterien die Validität (Gültigkeit), die Reliabilität (Zuverlässigkeit) und die Objektivität für Tests fest. Für summative Prüfungen werden des Weiteren noch Chancengerechtigkeit, Fairness und Ökonomie als Gütekriterien angeführt (Stieler, 2011, S. 25). In der Zusammenstellung von Prüfungsfragen liegt die Herausforderung darin, valide Fragen zu generieren, die abfragen, was in den Lehrzielen festgelegt ist. Objektivität und Reliabilität sind die Voraussetzung für eine valide Prüfung. Prüfungen, bei denen die Durchführung und Auswertung computerunterstützt realisiert werden, entsprechen dem Anspruch der Objektivität, weil die Bewertung nicht von intersubjektiven Einflüssen bestimmt wird. Die Durchführungsobjektivität der Prüfung kann durch die Auslagerung der Prüfungsaufsicht auf „fachfremde“ Personen, die keine inhaltlichen Fragen beantworten, gesichert werden.

Bei E-Assessment-Systemen gibt es in der Zwischenzeit eine Reihe von verfügbaren Frage- und Aufgabenformen, die – wie Single-Choice, Multiple-Choice, Ja / Nein bzw. Richtig / Falsch Zuordnungen – automatisch ausgewertet werden können. Daneben finden sich Matrix-Aufgaben, Lückentext- und Kurztextantworten sowie die Möglichkeit der Einbindung von Medien wie Bilder und Animationen. Je nach Fachgebiet stehen angepasste Aufgabenformen zur Verfügung, beispielsweise der Fragetyp *Buchungssatz* für das Fach Buchhaltung und Bilanzierung oder aber auch die Form eines *Open-Book-Exams*, bei dem die Lernenden Rechercheaufgaben in Form von Webquests lösen.

Bei der Betrachtung von E-Assessment-Systemen müssen mehrere Dimensionen berücksichtigt werden, um einen adäquaten und qualitativ hochwertigen Einsatz zu gewährleisten. Die technische Lösung stellt die Infrastruktur dar, auf der aufbauend didaktische, methodische und organisatorische Lösungen entwickelt werden können (Gruttmann, 2008, S. 17). Im vorliegenden Beitrag liegt das Hauptaugenmerk auf technischen, organisatorischen und rechtlichen Lösungen. Didaktische und methodische Überlegungen wurden bei der Entwicklung des Konzeptes für den elektronischen Prüfungsraum miteinbezogen, sie sind aber nicht Schwerpunkt dieses Beitrages.

Die Vorbereitung, Durchführung und Nachbereitung von Prüfungen erfordert eine Reihe von Aufgaben, die durch organisatorische und technische Maßnahmen unterstützt werden können. Das Konzept des hier vorgestellten elektronischen Prüfungsraumes unterstützt die Organisation von Prüfungen in unterschiedlichem Maße, wobei die räumlich-technische Ausstattung in den Mittelpunkt der Betrachtung gestellt wird.

Tabelle 1: Organisatorische Prozesse von elektronischen Prüfungen (in Anlehnung an Ehlers, 2013, S. 234)

Vorbereitung	Durchführung	Nachbereitung
Fragenkatalog erstellen und bereitstellen	Systembetrieb sicherstellen	Automatische / manuelle Korrekturen koordinieren und durchführen
Prüfung generieren	Anwesenheit / Identität der Teilnehmenden prüfen	Feedback an Lernende geben
Anwesenheitslisten erzeugen	Rechner- und Prüfungszugang koordinieren	Notenlisten generieren und an Studienadministration übermitteln
Raum-/ Zeitressourcen koordinieren	Organisatorische Fragen klären	Einsicht- und Einspruchnahme koordinieren
Studierenden- und Organisationsdaten in Prüfungssystem importieren	Manipulation / Täuschung verhindern	Archivierung sichern
Teilnehmende mit dem Prüfungssystem vertraut machen		

Neben der Unterstützung der organisatorischen Prozesse rund um eine hohe Anzahl an Prüfungen und der automatisierten Korrektur der Aufgabenstellungen, die vor allem Lehrenden und Prüfungsadministratoren zu Gute kommen, bieten E-Assessment-Systeme auch für Studierende eine Reihe von Vorteilen wie rasche Korrektur und Prüfungsergebnisse, Fairness und objektive Bewertung, Feedback und Prüfungseinsicht gleich im Anschluss an die Prüfung oder Fragenpools mit Musterklausuren.

Die Bedenken, die gegenüber der elektronischen Prüfung geäußert werden, beziehen sich einerseits auf didaktisch-methodische Aspekte – Prüfung von in erster Linie deklarativem Wissen – und andererseits auf organisatorisch-rechtliche Fragen wie Aufwand bei der Erstellung der Aufgaben, Zuverlässigkeit des Systems oder Fairness gegenüber Studierenden mit mangelnder Medienkompetenz.

Technische Umsetzung

Einrichtung Prüfungsraum

Für die elektronischen Prüfungen wurde ein PC-Labor mit 62 Arbeitsplätzen (zwei davon mit höhenverstellbaren Tischen für Rollstuhlfahrer) und einem Moderatorenarbeitsplatz eingerichtet, das ausschließlich für Prüfungen verwendet wird. Im Gegensatz zur üblichen Vorgangsweise, die Prüfungsarbeitsplätze mit Trennwänden vorne und an der Seite „einzuhausen“, wurde eine alternative Form gewählt, wie sie von Huth vorgestellt (2010) wurde. Die Bildschirme sind in die Tische eingelassen (siehe Abbildung 1); die 20 cm hohen Trennwände sollen insbesondere verhindern, dass unerlaubte Materialien auf den Tischen verschoben werden und sie unterbinden auch die Einsicht auf den rechten Arbeitsplatz. Um die Einsicht für größere Studierende auf den links gelegenen Bildschirm zu erschweren, wurden zusätzlich Polarisationsfilter auf die Bildschirme aufgebracht. Diese Lösung hat den Vorteil, dass die Studierenden durch wenige Aufsichtspersonen überblickbar sind, während sie sich sonst hinter ihrem jeweiligen Tisch nahezu vollständig verstecken können.

Der PC selbst ist in einem versperrten Schrank untergebracht. Das soll sowohl das versehentliche Aus- und Einschalten als auch das Verwenden eigener Devices wie USB-Sticks verhindern (siehe Abbildung 1).



Abbildung 1: Ausstattung Prüfungsraum

Der Moderatorenarbeitsplatz dient dazu, die einzelnen Prüfungsrechner via Netboot einzuschalten, dieses Ein- und Ausschalten kann auch auf einzelne Rechner bzw. Rechnerreihen beschränkt werden. Damit kann die Prüfungsaufsicht einen Sitzplan vorgeben, indem sie nur bestimmte Geräte für die Prüfung freischaltet.



Abbildung 2: Prüfungsraum – Übersicht Moderationsarbeitsplatz

Die Prüfungsarbeitsplätze sind mit einer Chipkartentastatur ausgestattet (siehe Abbildung 1) und alle Studierenden der Universität besitzen als Studierendenausweis eine entsprechend codierte Chipkarte (KeplerCard), mit der sie sich am jeweiligen Arbeitsplatz anmelden. Eine erfolgreiche Anmeldung wird auf dem Moderatorenarbeitsplatz registriert und durch die Matrikelnummer und das Foto vom Studierendenausweis angezeigt. Des Weiteren kann der Bildschirminhalt eines Rechners an den Moderatorenplatz geholt werden (siehe Abbildung 2).

Der gesamte Prüfungsraum ist auf möglichst hohe Autarkie ausgelegt. Es stehen daher alle erforderlichen Komponenten in unmittelbarer Nähe und sind entsprechend gegen Ausfall abgesichert. Für eine Standardprüfung besteht auch keine Verbindung zum Internet. Auf den Servern läuft als Anwendungssystem für die Prüfungen eine eigene Moodle-Instanz (Moodle 2.5), die durch ein selbst erstelltes Modul zur Integration der Bürgerkartensoftware erweitert wurde. Mit diesem Modul ist es möglich, auf die Signaturerstellungsdienste der Bürgerkartenumgebung nach dem österreichischen E-Government-Standard (A-SIT) zuzugreifen. Zusätzlich

wird für Lehrveranstaltungen in Rechnungswesen der zusätzliche Fragentyp Buchungssatz verwendet.

Die PCs laufen unter dem Betriebssystem LINUX mit abgespecktem Kernel und einem Browser. Die Entscheidung für LINUX erfolgte, weil damit ein schlankeres System konfiguriert werden konnte, aus dem viele Missbrauchsmöglichkeiten entfernt wurden. Bei einer Windows-Installation wäre das nicht möglich gewesen.

Organisatorisch-rechtliche Überlegungen

Maßgebliches Ziel bei der Einrichtung des Prüfungsraumes war, den Studierenden mehr Prüfungstermine bzw. kürzere Wiederholungszyklen anzubieten und so zu einer Studienbeschleunigung beizutragen. Dies ist allerdings nur dann realisierbar, wenn diese zusätzlichen Prüfungstermine ohne zusätzliche Belastung der Prüfenden angeboten werden können. Die Prüfungsaufsicht übernehmen Personen aus dem administrativen Bereich der Universität, für etwaige IT-technische Fragen ist immer auch ein Techniker bei den Prüfungen anwesend bzw. erreichbar. Bei üblichen schriftlichen Prüfungen treffen die Prüfer teilweise Entscheidungen ad hoc und klären auf diese Weise auftretende Problemfälle. Wenn mit der Anwesenheit solch entscheidungsbefugter Personen nicht gerechnet werden kann, sind detaillierte schriftliche Regelungen erforderlich, damit unklare Situationen möglichst vermieden werden. Für den Betrieb eines derartigen Prüfungsraumes sind daher umfangreichere organisatorische Regelungen erforderlich als sonst bei schriftlichen Prüfungen mit Fachaufsicht.

Zur Prüfung antreten darf nur, wer sich im Moodle-Prüfungskurs zur Teilnahme angemeldet hat und dessen Anmeldung gültig gespeichert ist. Das setzt eine aktive Zulassung zum Studium voraus und gegebenenfalls die Erfüllung bestimmter Voraussetzungen in einzelnen Lehrveranstaltungen. Beide Voraussetzungen werden elektronisch überprüft: Durch das Einlesen der Matrikelnummer aus der Chipkarte und dem Ablaufdatum wird das aktive Studium nachgewiesen und durch die erfolgreiche Suche dieser Matrikelnummer im Moodle-Kurs die spezifischen Voraussetzungen zu dieser Lehrveranstaltung geprüft.

Vor der eigentlichen Prüfung wird des Weiteren noch ein Dokument generiert und dem Studierenden zur digitalen Signatur angezeigt, welches folgende wesentliche Informationsblöcke enthält:

1. Aktuelle Aufnahme aus der Webcam des Prüfungsrechners, die ein Foto der Person, die tatsächlich am Computer arbeitet, zeigt.
2. Alle Stammdaten des Studierenden, einschließlich des Fotos auf dem Studierendenausweis auf Grund der aus der Chipkarte ausgelesenen Matrikelnummer
3. Besondere Prüfungsbedingungen, wie etwa zugelassene Prüfungsmittel.

Dieses Dokument wird digital signiert; dazu muss auf der KeplerCard die Bürgerkartensoftware installiert sein und der Studierende muss seine zum Installationszeitpunkt eingerichtete PIN eingeben. Die Signatur dieses Dokuments wäre rechtlich nicht notwendig, dieser Schritt dient nur dazu, vor der Prüfung sicherzustellen, dass die Studierenden die erforderlichen Vorbereitungen getroffen bzw. das erforderliche Wissen tatsächlich haben. Die Erfahrung hat nämlich gelehrt, dass der Hinweis auf diese Erfordernisse nicht ausreicht und immer wieder Personen mit mangelhafter Konfiguration bzw. vergessener PIN an der Abgabe der signierten Prüfung scheitern.

Grundsätzlich sind im Prüfungsraum keine Hilfsmittel erlaubt; die Studierenden haben daher unmittelbar im Eingangsbereich Schließfächer zur Verfügung, in denen sie alle Unterlagen, Notebooks, Handys und sonstige zur Kommunikation und Speicherung von Informationen geeigneten Geräte verschließen können.

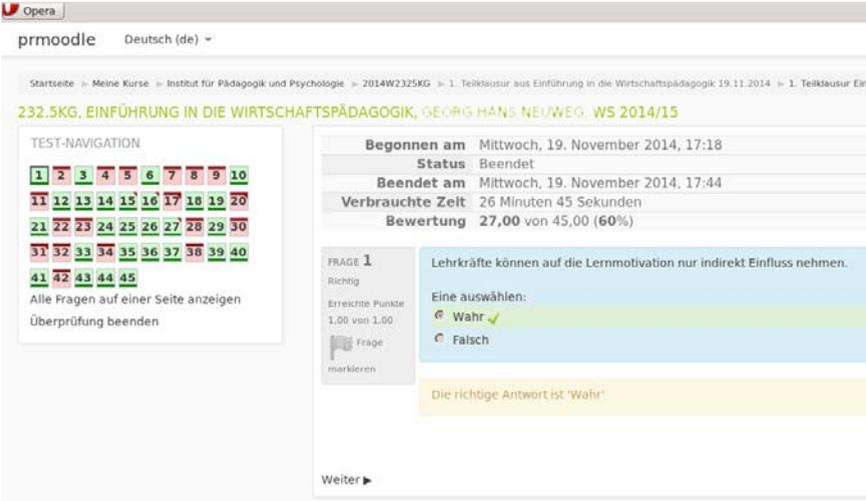
Nach dem erfolgreichen Abschluss der Unterfertigung der Prüfungsbedingungen startet der eigentliche Prüfungsvorgang. Studierende, die freiwillig oder auf Grund von technischen Rückmeldungen den Prüfungsraum vorher verlassen, wird dies nicht als Prüfungsantritt gewertet. Wie die Prüfung zusammengestellt wird, obliegt ausschließlich der Zuständigkeit der Prüfenden. Es stehen alle im Moodle verfügbaren Fragetypen einschließlich offener Fragen zur Verfügung.

Die Prüfung endet entweder durch die Abgabeentscheidung des Kandidaten oder durch Zeitablauf. In beiden Fällen wird jetzt nicht die Moodle Standardfunktion verwendet, sondern ein eigens programmierter Zwischenschritt eingefügt: die gestellten Fragen und die vom Studierenden gewählte(n) Antwortoption(en) bzw. seine Textantworten werden in einem Secure Viewer Modus angezeigt. Der Studierende hat das Dokument zu kontrollieren und mit der digitalen Unterschrift zu bestätigen (siehe Abbildung 3), dass dies seine gegebenen Antworten sind. Änderungen sind zu diesem Zeitpunkt nicht mehr möglich. Falls der Studierende behauptet, dass das Dokument fehlerhaft ist, kann dies anhand von Aufzeichnungen nachgeprüft werden (siehe Abschnitt 3.3 Sicherheit).



Abbildung 3: Signatur der Prüfung nach der Abgabe

Die Weitergabe der Prüfungsergebnisse an die Lehrenden erfolgt ebenfalls elektronisch, wobei bei reinen Multiple-Choice-Prüfungen bereits eine Liste erzeugt werden kann, die in das Beurteilungssystem der Universität importiert werden kann. Für die Prüfung von Stichproben bzw. im Verdachtsfall stehen für die Dauer der Einspruchsfrist alle Dokumente, die bei der Prüfung erstellt wurden, den Prüfenden zur Verfügung.



The screenshot shows a Moodle assessment interface. At the top, it indicates the browser is Opera and the language is Deutsch (de). The navigation path is: Startseite > Meine Kurse > Institut für Pädagogik und Psychologie > 2014W2325KG > 1. Teilklausur aus Einführung in die Wirtschaftspädagogik 19.11.2014 > 1. Teilklausur Ein. The course title is **232.5KG, EINFÜHRUNG IN DIE WIRTSCHAFTSPÄDAGOGIK, GEORG HANS NEUWEG, WS 2014/15**.

TEST-NAVIGATION

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45					

Alle Fragen auf einer Seite anzeigen
Überprüfung beenden

Beginnen am Mittwoch, 19. November 2014, 17:18
Status Beendet
Beendet am Mittwoch, 19. November 2014, 17:44
Verbrauchte Zeit 26 Minuten 45 Sekunden
Bewertung 27,00 von 45,00 (60%)

FRAGE 1
Richtig
Erreichte Punkte 1,00 von 1,00
Frage markieren

Lehrkräfte können auf die Lernmotivation nur indirekt Einfluss nehmen.

Eine auswählen:
 Wahr ✓
 Falsch

Die richtige Antwort ist 'Wahr'

Weiter ▶

Abbildung 4: Prüfungseinsicht

Unter denselben Bedingungen werden dem Studierenden die Beurteilungen unmittelbar angezeigt (siehe Abbildung 4). Grundsätzlich wird bei Multiple-Choice-Prüfungen das Recht auf Prüfungseinsicht unmittelbar und abschließend eingeräumt. Da das sonst übliche Recht auf Fotokopien für MC-Fragen nicht gilt (§ 49 UG), ist die unmittelbare Information im Prüfungsraum sogar förderlich, diesen Geheimhaltungsanspruch umzusetzen.

Sicherheitsüberlegungen

Neben den bereits erwähnten Merkmalen der Infrastruktur, die ebenfalls schon unter dem Aspekt einer sicheren Prüfungsumgebung ausgewählt wurden, werden zusätzlich noch folgende Maßnahmen ergriffen:

- Regelmäßige Screenshots werden vom System aufgezeichnet, um die Arbeitsschritte bzw. um die angekreuzten Antwortelemente nachvollziehen zu können. Um möglichst keinen Teilschritt zu übersehen, hat sich ein Intervall von fünf Sekunden als sinnvoll herausgestellt.
- Falls jemand bei der Endabgabe einen Fehler bemerkt und die Ergebnisse nicht digital signiert, wird vom System ersatzweise diese Sammlung von Screenshots dem unsignierten Dokument angefügt und damit signalisiert, dass diese Prüfungsarbeit einer manuellen Beurteilung bedarf.

Fazit

Der elektronische Prüfungsraum wurde im Sommersemester 2011 eingerichtet und steht seit 2014 mit allen dargestellten Funktionen zur Verfügung. Schon vorher wurden prototypisch einzelne Lehrveranstaltungen mit geringeren Sicherheitsanforderungen bzw. höherem Personaleinsatz Tausende von Prüfungen abgehalten. Gerade in der Anfangsphase, vor allem mit dem neuen Fragetyp *Buchungssatz*, sind Fehler vorgekommen, die zu „Null-Punkte-Ergebnissen“ geführt haben, obwohl die Prüfungsaufsichten richtige Teilantworten wahrgenommen haben. Erfahrungen wie diese liegen den heute implementierten Sicherheitsfunktionen zu Grunde. Und wenn manches heute überschießend erscheinen mag, so liegt für jede dieser Maßnahmen eine entsprechende Erfahrung aus der ersten Phase der elektronischen Prüfungen zugrunde.

Von Studierendenseite wird insbesondere die rasche Information über die Beurteilung und die unverzügliche Einsicht, welche der Fragen richtig oder falsch beantwortet wurde, geschätzt.

Zur Mitte und zum Ende des Semesters ist der Prüfungsraum beinahe an den Kapazitätsgrenzen, ohne dass dafür Werbemaßnahmen ergriffen oder eine offizielle Empfehlung gegeben wurde. So wurden allein im Wintersemester 2014 / 15 1800 Prüfungen in 60 Prüfungsdurchgängen abgenommen, wobei im Schnitt um ein Viertel mehr Studierende zu den Prüfungen angemeldet waren. Die positive Empfehlung von Kolleginnen und Kollegen hat sich auch hier als effiziente Werbemaßnahme bestätigt.

Referenzen

A-SIT Zentrum für sichere Informationstechnologie - Austria (Hrsg.). Abgerufen unter <http://www.buergerkarte.at/downloads-karte.html> [30.04.2015]

Achtenhagen, F. & Beathge, M. (2007). Kompetenzdiagnostik als Large-Scale-Assessment. *Zeitschrift für Erziehungswissenschaft, Sonderheft 8/2007*, 51-70.

Bergmann, J. & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Eugene: ISTE.

Ehlers, J. P., Gütl, C., Höntzsch, S., Usener, S. & Gruttmann, S. (2013). Prüfen mit Computer und Internet: Didaktik, Methodik und Organisation von E-Assement. In M. Ebner & S. Schön (Hrsg.), *L3T: Lehrbuch für Lernen und Lehren mit Technologien* (2. Aufl., S. 228-237), Abgerufen unter <http://creativecommons.org/licenses/by-sa/3.0/de> [30.04.2015]

Eilers, B., Gruttmann, S. & Kuchen, H. (2008). Konzeption eines integrierbaren Systems zur computergestützten Lernfortschrittskontrolle. In H. L. Grob, J. vom Brocke & C. Buddendick (Hrsg.), *E-Learning-Management* (S. 213-232). München: Vahlen.

Gruttmann, S. (2008). *Formatives E-Assessment in der Hochschullehre: Computerunterstützte Lernfortschrittskontrollen im Informatikstudium*. Münster: Monsenstein und Vannerdat.

Huth, D. (2010). *E-Assessment an der Universität Duisburg-Essen im „Kompetenzzentrum für PC-gestützte Prüfungen“* [Präsentationsfolien, 2. Symposium E-Learning an Hochschulen, TU Dresden, 9. März 2010, Folie 13]. Abgerufen unter https://www.uni-due.de/imperia/md/content/zim/projekte/e-assessment_ude_09032010_v2.pdf [30.04.2015]

Müller, A. (2011). Prüfungen als Lernchance: Vorstellung und Überprüfung eines Rahmenmodells. *Zeitschrift für Hochschulentwicklung ZFHE*, 6(1), 72-94.

Reinmann, G. (2011). Kompetenz – Qualität – Assessment: Hintergrundfolie für das technologiebasierte Lernen. In M. Mühlhäuser, W. Sesink & A. Kaminski (Hrsg.), *Interdisziplinäre Zugänge zum technologiegestützten Lernen* (S. 467-493). Münster: Waxmann.

Stieler, F. (2011). *Validität summativer Prüfungen: Überlegungen zur Gestaltung von Klausuren*. Bielefeld: Janus Presse.

Van Treeck, T., Himpsl-Gutermann, K. & Robes, J. (2013). Offene und partizipative Lernkonzepte: E-Portfolios, MOOCs und Flipped Classrooms. In M. Ebner & S. Schön (Hrsg.), *L3T: Lehrbuch für Lernen und Lehren mit Technologien*. (2. Aufl., S. 287-300). Abgerufen unter <http://l3t.tugraz.at/index.php/LehrbuchEbner10/article/view/149> [30.04.2015]

Wesiak, G., AL-Smadi, M., Höfler, M. & Gütl, C. (2013). Assessment for complex learning resources: Development and validation of an integrated model - Special focus paper, *International Journal of Emerging Technologies in Learning*, 8, 52-61.

Vita

Mag. Dr. Elisabeth Katzlinger ist Assistenzprofessorin am Institut für Datenverarbeitung in den Sozial- und Wirtschaftswissenschaften der Johannes Kepler Universität Linz, Österreich. Sie studierte Betriebswirtschaft und Wirtschaftspädagogik an der Universität Linz. Ihre Forschungsschwerpunkte sind zurzeit: E-Learning, Game-Based Learning sowie E-Tutoring. Die Lehrtätigkeit umfasst Business und Internet sowie Digital Business Management. Weitere Arbeitsbereiche sind: MUSSS - Multimedia Studienservices in den Sozial- und Wirtschaftswissenschaften sowie die

Leitung des Vorbereitungslehrgangs für Studienberechtigungsprüfung (Studium ohne Matura).

A. Univ.-Prof. DDr. Johann Höller ist Institutsvorstand des Instituts für Datenverarbeitung in den Sozial- und Wirtschaftswissenschaften sowie Vorsitzender der Studienkommissionen Webwissenschaften und Digital Business Management (beides Masterstudien) an der Johannes Kepler Universität Linz, Österreich. Des Weiteren leitet er MUSSS - Multimedia Studienservices in den Sozial- und Wirtschaftswissenschaften. Er studierte Rechtswissenschaften und Betriebswirtschaft. Er ist Professor für Informationsverarbeitung und Informationsrecht. Die Forschungsinteressen liegen auf Digital Business und Informationsrecht.

Qualitätssicherung bei Prüfungen

Qualitätssicherung elektronischer Prüfungen mit geschlossenen Aufgabenformaten

Zusammenfassung

Elektronische Prüfungen mit geschlossenen Aufgabenformaten (E-Klausuren) bieten die Möglichkeit detailliert abzubilden, welche Learning Outcomes eines Moduls von den Studierenden in welchem Maße erreicht wurden. Dieses Ziel wird allerdings nur erreicht, wenn der gesamte Prüfungsprozess durch eine ausreichende Qualitätssicherung begleitet und kontrolliert wird.

Grundlage objektiver, valider und reliabler Prüfungen und damit die erste Dimension von Qualitätssicherung bildet die Kohärenz von Learning Outcomes, Prüfungs- und Lernprozess, welche durch Constructive Alignment in der hochschuldidaktischen Planung vorbereitet und testtheoretisch gesichert werden kann. In einer zweiten Dimension werden die zeitlichen Phasen der Entwicklung, Durchführung, Auswertung und Evaluation einer Prüfung mit ihren typischen Anforderungen in der Qualitätssicherung berücksichtigt. Wichtige Einzelfragen in diesem Prozess sind die Entwicklung der Prüfungssitems im Prozess des Constructive Alignments, ein mindestens zweistufiges Itemreview, die Konstruktion des Prüfungsmodells (Blueprint), die Analyse der Prüfungsergebnisse im Hinblick auf geeignete statistische Kennwerte und die Entwicklung eines Bewertungsmodells, in das die Ergebnisse der Evaluation einfließen können und das die Festlegung der Bestehensgrenze reflektiert und methodisch absichert.

Qualitätssicherung: Mehr als ein technisches Problem

Qualitätssicherung von Prüfungen in der akademischen Praxis

Seit rund zwanzig Jahren gehört das Thema Qualitätssicherung im Bildungswesen zu den meistdiskutierten, sowohl im schulischen Bereich (KMK, 1997) als auch im Hochschulbereich (Beise, Hungermann & Wannemacher, 2014). Der Qualitätssicherung von Hochschulprüfungen kommt in diesem Zusammenhang eine besondere Bedeutung zu, weil sie sich dem Aspekt der Zielerreichung akademischer Lernprozesse widmet. Prüfungen mit geschlossenen Aufgabenformaten haben vielfach mit dem Vorurteil zu kämpfen, dass sie lediglich eine billige Variante anderer, ver-

meintlich „an sich“ hochwertiger akademischer Prüfungsformate (Prüfungsgespräch, Essay-Klausur, Freitextaufgaben) darstellten. Die Auseinandersetzung mit dem Thema Prüfungsqualität darf sich daher nicht auf die technische Ebene und die damit verbundenen Fragen wie zum Beispiel die der Verfahrenssicherheit, der Prüfungssoftware oder der testtheoretischen Gütekriterien beschränken, sondern muss sehr viel fundamentalere Aspekte beachten (Tinnefeld, 2013).

Qualitätssicherung wird in diesem Beitrag begriffen als ein Teil des umfassenderen Qualitätsmanagements. Im Kontext dieses Qualitätsmanagements geht es nicht nur um Effizienz-, sondern auch um Effektivitätsziele. Effizienz zielt darauf,

- die Dinge richtig zu tun und
- die Dinge zur richtigen Zeit zu tun.

Effektivität hingegen zielt weiter gehend darauf,

- die richtigen Dinge zu tun.

Die Beurteilung der Qualität akademischer Prüfungen wird in der Praxis durch zwei Momente erheblich erschwert: Erstens gibt es tradierte Prüfungsformate, die teils fächerspezifisch sind und in jedem Falle die tradierte Prüfungskultur des Faches repräsentieren. Zweitens werden Hochschullehrer nach Gesichtspunkten rekrutiert, in denen die akademische Lehre im Allgemeinen und die Gestaltung von Prüfungen im Besonderen von untergeordneter Bedeutung ist. Für den „handwerklichen“ Teil dieses Berufsfeldes gibt es inzwischen zwar hochschuldidaktische Qualifikationsprogramme, aber keine curricularisierte Ausbildung. In der Folge ist es nicht verwunderlich, dass sich die Betroffenen in der Regel an der tradierten Prüfungskultur ihres Faches orientieren, ohne diese in Frage zu stellen.

In diesem Sinne kann man Qualitätsentwicklung als einen durch Reflexion der gegenwärtigen Praxis angestoßenen Optimierungszirkel (Abbildung 1) verstehen:



Abbildung 1: Optimierungszyklus der Qualitätsentwicklung

Bereits die Analyse der Ausgangssituation führt in der Regel dazu, die bisher vielleicht nicht hinterfragte Prüfungs-tradition in ihren Vorzügen und ihren Schwachstellen zu verstehen, Bewusstsein zu schaffen, Friktionen in Verfahrens-abläufen zu erkennen und verbesserte Modelle zu entwickeln.

Zu den klassischen Gütekriterien eines Tests (Objektivität, Validität und Reliabilität) ist im Hinblick auf traditionelle Prüfungsformate (mündliche Prüfung, Essay-Klausur) viel Skeptisches vorgetragen worden, das an dieser Stelle nicht wiederholt werden soll (Ingenkamp, 1971). Im folgenden Beitrag konzentrieren wir uns vielmehr auf grundsätzliche Erwägungen zur Verbesserung von Validität und Reliabilität schriftlicher Prüfungen mit geschlossenen Aufgabenformaten im Sinne eines Erfahrungsberichtes aus zehnjähriger Prüfungspraxis.

Rahmenbedingungen, Validität und Reliabilität

Die richtigen Dinge zur richtigen Zeit in der richtigen Art und Weise zu tun, ist auch im Kontext von Hochschulprüfungen an kommunikative Vergewisserung gebunden. Für die ontogenetische Entwicklung des Leistungsmotivs gilt als gut gesichert (Heckhausen, 1980; McClelland et al., 1953), dass die Beurteilung einer Handlung als Leistung an die Existenz eines sozial anerkannten Gütemaßstabs gebunden ist, der aus der Sache selbst resultieren oder sozial vermittelt sein kann. Für Hochschulprüfungen als Modus einer Leistungsmessung bedeutet dies insbesondere, dass diese Gütemaßstäbe unter Experten ausgehandelt und mit zu prüfenden Novizen kommuniziert werden müssen, wenn akademische Prüfungen nicht nur den Status eines Initiationsritus haben sollen. Die Frage, ob wir in einer Prüfung

die richtigen Dinge tun, betrifft im Kern die Validität einer Prüfung. Für die Validität kommen unterschiedliche Aspekte zum Tragen:

- Die *Inhaltsvalidität* fragt danach, ob die Messung ein Konstrukt in all seinen Inhaltsbereichen erfasst. Im Analogieschluss bedeutet das: Erfasst eine Prüfung das zu messende Konstrukt, d. h. die Zielkompetenzen, in all seinen Aspekten. Inhaltsvalidität betrifft bei Prüfungen also die Frage, ob und wie gut die inhaltliche Operationalisierung von Zielkompetenzen durch die Prüfungsfragen gelingt. Inhaltsvalidität lässt sich unserer Erfahrung nach am besten durch Expertenrating erfassen und beschreiben.
- Bezogen auf die Beurteilung von Tests kann man das Thema *Konstruktvalidität* „rückwärts“ lesen und fragen, ob die intendierten Kompetenzziele durch die vorliegenden Tests erfasst werden. Dazu lassen sich *konvergente* und *diskriminante (divergente) Validität* bestimmen über die Prüfung, ob Messdaten von Tests, welche dieselbe (bzw. verschiedene) Kompetenz abbilden, erwartungsgemäß hoch (bzw. niedrig) korrelieren.
- Schließlich ist nach der *Kriteriumsvalidität* von Prüfungen als Test zu fragen und damit nach dem Zusammenhang des Messinstruments mit empirischen Kriterien, sowohl als *diagnostische Validität* und, im Hinblick auf die Erweisbarkeit der Kompetenzziele, mehr noch als *prognostische Validität*.

Zu diesen Aspekten gibt es derzeit keine gesicherte Prüfungstheorie, auf die sich die Entwicklung geeigneter Messverfahren stützen könnte: Die genannten Aspekte stellen indessen Prüfungsaspekte dar, die in den Reflexionshorizont der Prüfungskonstruktion mit einbezogen werden sollten.

Man kann ferner die Frage aufwerfen, ob die Grundidee der Testkonstruktion im Rahmen der klassischen Testtheorie, auf die sich ja die Gütemaße Objektivität, Validität und Reliabilität beziehen, auf die Situation des akademischen Prüfens übertragbar ist. Tests und Testtheorie entstammen dem Bereich der (psychologischen) Diagnostik. Hier sind zwei Phasen zu unterscheiden: die Testkonstruktion und die Testanwendung. Während der Testkonstruktion werden die Gütemaße beobachtet und zur qualitativen Weiterentwicklung des Tests benutzt. Wenn der Test fertig entwickelt ist, wird er angewandt: Dann werden die Testergebnisse nicht auf Trennschärfe und Schwierigkeit hin analysiert. Es spricht manches dafür, dass Prüfungen nicht mit diagnostischen Tests gleichgesetzt werden können – Klassenarbeiten in der Schule sind etwas anderes als genormte Schulleistungstests. Und insofern ist die Frage zu stellen, ob die Itemanalyse, die man im Bereich von E-Assessments in Analogie zur psychologischen Diagnostik durchführt, in der Sache gerechtfertigt ist. Allerdings verwenden wir einige empirische Indikatoren wie Schwierigkeit und Trennschärfe, um möglichen Schwachstellen unserer Prüfungskonstruktion auf die Spur zu kommen.

Qualitätssicherung im Modified Constructive Alignment

Constructive Alignment gehört zu den Schlüsselbegriffen in der aktuellen Hochschuldidaktik, durch die eine vielerorts als Paradigmenwechsel apostrophierte „Shift from teaching to learning“ markiert wird (Biggs & Tang, 2011). „Beim Constructive Alignment Konzept geht es im Kern darum, dass die intendierten Outcomes des Lernprozesses klar definiert und den Studierenden explizit verdeutlicht werden und die Prüfungs- und Lernaktivitäten stringent auf die Learning Outcomes abgestimmt werden“ (Schaper 2012, S. 62). Unter Learning Outcomes versteht man die intendierten Ziele: „Learning outcomes describe what a learner is expected to know, understand and be able to do after successful completion of a process of learning“ (Europäische Gemeinschaft 2009, S. 11). Dabei ist darauf hinzuweisen, dass sich das Prinzip Constructive Alignment auf gesamte Studiengänge und auf ganze Module bezieht. Oftmals besteht aber die Herausforderung darin, dass für bereits bestehende Module qualitätsgesicherte Prüfungen entwickelt werden sollen, ohne dass es in der Regel möglich wäre, die Kompetenzziele dieser Module zu verändern. Diese Modifikation bezeichnen wir als Modified Constructive Alignment (MCA).

Constructive Alignment, auch in seiner modifizierten Form, lässt sich in drei Planungsschritten konkretisieren, die in Abbildung 2 zusammengefasst sind: Zentral ist dabei die Ausrichtung der Prüfung an den Learning Outcomes und die anschließende Orientierung der zu arrangierenden Lernsituationen an den festgelegten Prüfungsformaten und -inhalten. Dieses hochschuldidaktische Planungsmodell stellt die erste Ebene der Qualitätssicherung von Prüfungen dar, auf der zweiten Ebene sind die Prozessschritte zur Entwicklung und Durchführung der Prüfung zu sichern. Diese bilden den eingangs erwähnten Optimierungszyklus der Qualitätssicherung ab und umfassen die Entwicklung, die Durchführung, die Auswertung und Evaluation und die Überarbeitung der Prüfung, in unserem Falle der E-Klausur (Abbildung 3).

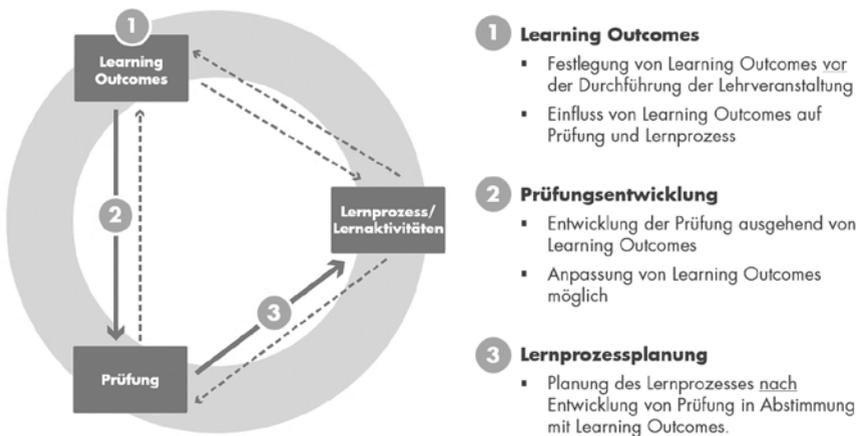


Abbildung 2: Planungs Schritte im Constructive Alignment

In der Praxis hat es sich bewährt, darüber hinaus die Qualitätssicherung einer E-Klausur sowohl auf der Ebene des gesamten Tests als auch auf der Ebene der einzelnen Aufgaben (Items) zu betreiben. Auf der Ebene des gesamten Tests ist die Erfüllung der klassischen Gütekriterien Objektivität, Validität und Reliabilität zu fordern. Objektivität im Sinne von Durchführungsobjektivität und Auswertungsobjektivität kann man bei E-Klausuren als gegeben ansehen. Die Validität der Prüfung bezieht sich bei dieser Prüfungsform in erster Linie auf die Frage der *inhaltlichen* Gültigkeit. Sie ist zu sichern über Experten-Rating hinsichtlich der Fragen und über einen Blueprint der E-Klausur, verstanden als Strukturplan der Prüfung hinsichtlich der abgebildeten Inhalte, der Aufgabenformate und der Anforderungsstufen der Items. Die Reliabilität betrifft die innere Konsistenz der Prüfung und lässt sich im Rahmen der Klassischen Testtheorie (KTT) mit den bekannten Mitteln der Konsistenzanalyse bestimmen. Auf der Ebene der einzelnen Items erfolgt die Qualitätssicherung im Rahmen der KTT durch die Überprüfung der empirischen Schwierigkeit ($20 \leq P_i \leq 80$) und der Trennschärfekoeffizienten ($r \geq .2$). Problematisch ist in diesem Zusammenhang, dass diese Werte stichprobenabhängig sind und dass sie demzufolge für jede Aufgabe von Prüfungsgruppe zu Prüfungsgruppe schwanken. Einen Ausweg böte hier allein die Anwendung der probabilistischen Testtheorie (Kubinger, 2009, 2014).

Zum Prozess der Qualitätssicherung im Modified Constructive Alignment

Die Prozesselemente der Qualitätssicherung von E-Klausuren im Rahmen des Constructive Alignments sind in Abbildung 3 im Überblick dargestellt. Im Folgenden werden die einzelnen Prozessschritte näher ausgeführt.

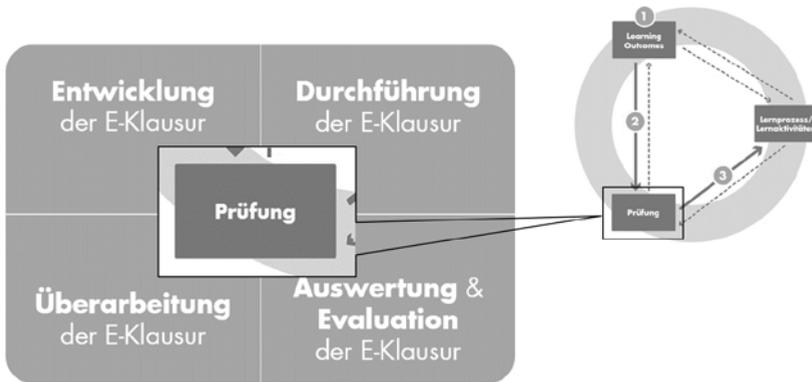


Abbildung 3: Prozessmodell für E-Klausuren

Entwicklung der E-Klausur

In der Phase der Entwicklung ist sicherzustellen, dass ausreichende Kenntnisse über die Grundprinzipien, im vorliegenden Fall des Constructive Alignments (Biggs & Tang 2011) und der Taxonomie nach Anderson und Krathwohl (2001) innerhalb des Kollegiums vorhanden sind. Zweitens ist die Gruppe zu schulen hinsichtlich der inhaltlichen und formalen Normen geschlossener Aufgaben. Es hat sich als hilfreich erwiesen, verschiedene Arbeitshilfen zur Verfügung zu stellen, die bei der Einhaltung dieser Normen unterstützen, z. B. indem man Schreibvorlagen entwickelt, in denen diese Vorgaben zur Gestaltung und zur Einordnung in die Taxonomie enthalten sind. Ein weiterer wichtiger Punkt betrifft in der Entwicklungsphase die Erstellung des Blueprints der E-Klausur, in dem die Anzahl, die Anforderungsstufe, die Wissensdimension, die Inhaltsbereiche der Items festgelegt werden. Schließlich gehört die Ermittlung der aufgabenspezifischen Ratewahrscheinlichkei-

ten in die Entwicklungsphase der Klausur, die in die Erstellung des Bewertungsmodells einfließen sollen.

Bei der Erstellung von geschlossenen Aufgaben ist bekanntlich sicherzustellen, dass mindestens zwei Prüfer jedes Item hinsichtlich seiner Tauglichkeit beurteilen. Es hat sich bewährt, den Prozess der Aufgabenerstellung durch die Verwendung von Arbeitshilfen zu organisieren, die wesentliche Elemente der Qualitätssicherung beinhalten. Darüber hinaus empfiehlt es sich vor dem Hintergrund des Zweiprüfer-Prinzips, dass jede Aufgabe hinsichtlich ihrer formalen Richtigkeit, ihrer Formulierungsklarheit und der Taxonomiestufe einem zweistufigen Reviewprozess unterworfen wird. In der ersten Stufe beurteilt ein zweiter Prüfer die erstellte Aufgabe, in der zweiten Stufe revidiert die gesamte Prüfergruppe die neu erstellten Aufgaben abschließend noch einmal. Dieser Reviewprozess, der ebenfalls durch Arbeitshilfen nach gleichen Kriterien organisiert wird, stellt ein zentrales Qualitätselement in der Entwicklungsphase dar: Hier werden mögliche Schwächen von Items zuverlässig erkannt und revidiert.

Durchführung der E-Klausur

Für die Durchführung der Klausur ist wichtig, dass die Studierenden mit der Benutzeroberfläche des Programms und mit der Prüfungsphilosophie vertraut sind. Dies lässt sich am besten erreichen, indem man ihnen eine Probeklausur zur Verfügung stellt, die in Aufbau und Umfang der echten Klausur entspricht, und die Korrektur der eigenen Antworten an die Teilnehmer zurückgibt.

Für die Akzeptanz des Prüfungsinstruments ist diese Transparenz von ausschlaggebender Bedeutung: Die Probeklausur nimmt Studierenden die Angst vor dem Unbekannten. Transparenz ist darüber hinaus in der Offenlegung des Bewertungsmodells geboten, z. B. in der Frage, ob bei Multiple-Choice-Fragen Punkte gesammelt werden können oder ob es einen Alles-oder-nichts-Bewertungsmodus gibt.

Drittens ist darauf zu achten, dass *jedes* Item eine genaue Instruktion für eine angemessene Aufgabebearbeitung enthält. Unter Prüfungsbedingungen muss mit einer für Studierende potenziell stressbehafteten Situation gerechnet werden, in der sich nichts „von selbst“ versteht, sondern durch exakte Instruktion in jedem Einzelfall begleitet werden muss.

Schließlich sei hier eher am Rand das technische Risikomanagement erwähnt. Dazu gehört die technische Überwachung des Prüfungservers, die Bereitstellung eines detaillierten Logfiles und die Dokumentation von inhaltlichen, formalen und technischen Problemen.

Auswertung und Evaluation

Nach der Durchführung der Klausur und ihrer automatisierten Auswertung sind die Ergebnisse auf Schlüssigkeit und Richtigkeit zu überprüfen. Hinweise auf mögliche Schwachstellen ergeben sich aus der Item-Analyse: Die empirische Schwierigkeit sollte zwischen 20 und 80 liegen, die Trennschärfe größer als 0.2 sein. In Abhängigkeit von diesen Ergebnissen kann man die Items erneut analysieren, um Hinweise auf mögliche Ursachen zu erhalten. Dabei ist es durchaus in das Ermessen des Prüfers gestellt, wie er mit den Ergebnissen umgeht: Eine Aufgabe mit einer Schwierigkeit von 10 wurde von 90% der Teilnehmer falsch gelöst. Ob das auf einen Fehler in der Aufgabenstellung zurückzuführen ist oder auf andere Faktoren, muss im Einzelfall gewürdigt werden. Falls sich Prüfer dazu entschließen, Aufgaben aus der Wertung zu nehmen, werden die individuell erreichten Punkte auf den Vergleichspunktwert renormiert. Dadurch wird sichergestellt, dass kein Teilnehmer durch den Ausschluss dieser Aufgabe aus der Wertung benachteiligt wird.

Auf der Ebene des gesamten Tests wird die innere Konsistenz durch die Berechnung von geeigneten Reliabilitätskoeffizienten ermittelt. Verbreitet ist die Benutzung von Cronbachs Alpha. Wir verwenden als Kenngröße Kuder-Richardson und empfehlen $0.8 \leq KR_{20} \leq 0.9$.

Überarbeitung der E-Klausur

Die Optimierung betrifft zwei Ebenen. Durch die Auswertung der Klausur und ihrer Korrektur lassen sich inhaltliche Kriterien für die Organisation der Lehrveranstaltung und die (Weiter-)Entwicklung der Klausur gewinnen: Sind die Learning Outcomes und die durch sie angesteuerten Kompetenzziele erreicht worden? Welche Prüfungsergebnisse deuten darauf hin, dass dies besonders gut bzw. weniger gut erreicht wurde? Welche Veränderung in den Lernarrangements können getroffen werden, um die intendierten Learning-Outcomes besser zu erreichen?

Auf einer zweiten Ebene werden Anhaltspunkte gewonnen, welche Innovationen des verwendeten Prüfungssystems notwendig oder wünschenswert sind: Gibt es Aufgabenformate, mit denen sich die intendierten Learning-Outcomes besser überprüfen lassen würden, die aber im Prüfungssystem nicht zur Verfügung stehen? Ist eine Korrektur-Historie im Bereich eventuell erforderlicher manueller Korrekturen verfügbar? Lassen sich Bewertungsoptionen in der vom Prüferkollegium gewünschten Form im Prüfungssystem darstellen?

Fazit und Ausblick

Zusammenfassend lässt sich sagen, dass die Qualitätssicherung von elektronischen Prüfungen mit geschlossenen Aufgabenformaten eine komplexe Anforderung ist, in

denen Elemente von technischer Kontrolle und sozialem Commitment eng miteinander verzahnt werden müssen. Für die klassischen Gütekriterien Objektivität, Validität und Reliabilität können Sicherungsverfahren angegeben werden, die sich derzeit aber nur zum kleineren Teil innerhalb einer Prüfungssoftware darstellen lassen: So ist man bei der Validität auf Experten-Ratings und Blueprints verwiesen, die sich nicht im Workflow der gängigen Prüfungssoftware darstellen lassen, sondern „externe“ Lösungen erfordern. Wünschenswert wäre, eine Softwareunterstützung für den gesamten Workflow zu haben. Ob dies in einem einzelnen Programm geschieht oder in separaten, aber gut integrierten Apps, sei dahin gestellt. Unbefriedigend sind die Instrumente der klassischen Testtheorie, weil Trennschärfe und Schwierigkeit stichprobenabhängig sind und daher keine guten Maße für die weitere Verwendbarkeit von Items generieren können. Hier könnte die Anwendung der Item-Response-Theorie (IRT) einen Schritt weiter führen. Bleibt man bei der klassischen Testtheorie, so sind für den Bereich von Mehrfachwahlaufgaben, in denen mehr als eine Antwortmöglichkeit richtig sein kann, mathematische Verfahren zu entwickeln, um eine Analogie zur Trennschärfe und Schwierigkeit zu bekommen, die ja nur für dichotome Items berechnet werden kann.

Ein schwerwiegendes Problem ist der Umgang mit der Ratewahrscheinlichkeit, weil sich in der Literatur Hinweise finden lassen, dass die Ratewahrscheinlichkeit sich mit der Anzahl der Items verändern könnte. Auch hier könnte sich die Anwendung spezieller Methoden der IRT als zielführend erweisen (Kubinger 2014, S. 170).

Referenzen

- Anderson, L. & Krathwohl, D. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Beise, A., Jungermann, I. & Wannemacher, K. (Hrsg.). (2014). *Qualitätssicherung von Studiengängen jenseits der Programmakkreditierung: Neue Herausforderungen für Hochschulsteuerung und Organisationsentwicklung (Forum Hochschule 1/2014)*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung.
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4. Aufl.). Maidenhead: McGraw Hill Education and Open University Press.
- Europäische Gemeinschaft (2009): ECTS-Leitfaden.
Abgerufen unter http://ec.europa.eu/education/tools/docs/ects-guide_de.pdf [16.05.2015]
- Heckhausen, H. (1980). *Motivation und Handeln*. Berlin: Springer.
- Ingenkamp, K. (1971). *Die Fragwürdigkeit der Zensurengebung*. Weinheim: Beltz.

KMK (1997). Grundsätzliche Überlegungen zu Leistungsvergleichen innerhalb der Bundesrepublik Deutschland: Konstanzer Beschluss (Beschluss der Kultusministerkonferenz vom 24.10.1997). Abgerufen unter <http://www.kmk.org/bildungschule/qualitaetssicherung-in-schulen.html> [16.05.2015]

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Göttingen: Hogrefe.

Kubinger, K. D. (2014). Gutachten zur Erstellung „gerichtsbarer“ Multiple-Choice-Prüfungsaufgaben. *Psychologische Rundschau*, 65(3), 169-178.

McClelland, D., Atkinson, J. W., Clark, R. A. & Lowell, E. L. (1953). *The achievement motive*. Princeton: Van Nostrand.

Schaper, N. (2012). Fachgutachten zur Kompetenzorientierung in Studium und Lehre. Abgerufen unter http://www.hrk-nexus.de/fileadmin/redaktion/hrk-nexus/07-Downloads/07-02-Publikationen/fachgutachten_kompetenzorientierung.pdf [16.05.2015]

Tinnefeld, T. (2013). *Dimensionen der Prüfungsdidaktik: Analysen und Reflexionen zur Leistungsbewertung in den modernen Fremdsprachen*. Saarbrücken: Hochschule für Technik und Wirtschaft des Saarlandes.

Abgerufen unter <https://drive.google.com/file/d/oB-2HUXwM3-xiYXhXQndnSTVCcDg/edit?usp=sharing> [10.09.2015]

Vita

Heinz-Werner Wollersheim, geb. 1957, Univ.-Prof., Dr. phil. habil., ist seit 1993 Inhaber des Lehrstuhls für Allgemeine Pädagogik an der Universität Leipzig. Von 1999 bis 2002 war er Sprecher des Sonderforschungsbereichs 417 der DFG an der Universität Leipzig. Seit 2004 umfangreiche Erfahrungen mit der Gestaltung und Durchführungen von E-Assessments mit großen Gruppen. Seit 2009 bis heute Studiengangverantwortlicher des Masterstudiengangs „Begabungsforschung und Kompetenzentwicklung – Studies in Abilities and Development of Competences.“ Arbeits- und Forschungsschwerpunkte: Begabungsforschung und Begabtenförderung, E-Learning und E-Assessment, Hochschuldidaktik, Historische Bildungsforschung.

Qualitätsmanagement bei E-Klausuren: Risikoanalyse und QS-Maßnahmen für Strukturen, Prozesse und Inhalte

Zusammenfassung

Der Beitrag zeigt vor dem Hintergrund der seit 2004 an der Universität Bremen vom Zentrum für Multimedia (ZMML) gesammelten Erfahrungen mit der Organisation und Durchführung von E-Klausuren eine Reihe von Qualitätssicherungsmaßnahmen auf den Ebenen Informationssicherheit, Prüfungsprozesse und Prüfungsinhalte. Eine wichtige Rolle spielt dabei das 2007 eröffnete Testcenter als infrastrukturelle Basis der E-Assessment-Services. Gegenüber der vorherigen PC-Pool-Lösung konnte die Durchführungssicherheit gesteigert, der Ressourceneinsatz pro Prüfung deutlich reduziert sowie Performanz und Vielfalt der Prüfungsszenarien erhöht werden. Neben dem Einsatz des professionellen Prüfungssystems der Firma LPLUS hat sich die im ZMML entwickelte Software eXamControl für die Unterstützung und Kontrolle der Prüfungsprozesse bewährt. Ein Ticket-System sorgt dafür, dass die prozessbegleitende Kommunikation an zentraler Stelle dokumentiert und verfolgt werden kann. Neben diesen technisch-organisatorischen Maßnahmen wird vom ZMML eine Reihe von begleitenden Serviceleistungen angeboten, die die Lehrenden bei der Steigerung der Prüfungsqualität unterstützen. Dazu zählen neben der prüfungsdidaktischen Beratung die feste Einbindung des ZMML in den Qualitätssicherungszyklus bei der Erstellung der Prüfungsfragen und die Bereitstellung von Fragentemplates und Checklisten.

Einleitung

Diagnostische und summative E-Assessments in Form von E-Klausuren, Eingangstests und Softwareprüfungen sowie Mischformen wie softwareintegrierte E-Klausuren werden seit 2004 im Rahmen eines regulären E-Learning-Dienstes des ZMML durchgeführt (eassessment.uni-bremen.de). 2007 wurde ein Testcenter eingerichtet, in dem aktuell pro Jahr ca. 16.000 Prüfungsleistungen in 160 Lehrveranstaltungen abgenommen werden (Stand Sommersemester 2014). Wie alle universitären Prüfungsformen, die im rechtlichen Sinne Einfluss auf die freie Berufswahl haben, stellen auch E-Klausuren hohe Ansprüche an das Qualitätsmanagement.

Jede E-Klausur ist ein sensibles Projekt mit einer unverrückbaren Deadline und dem Anspruch, dass zum Prüfungszeitpunkt technisch, organisatorisch und inhaltlich keine gravierenden Fehler auftreten. Neben Beratung, Schulung und Support sind somit Multiprojektmanagement und das in einen kontinuierlichen Verbesserungsprozess eingebettete Qualitäts- und Risikomanagement die wichtigsten Aufgaben der Servicestelle.

Qualität wird hier definiert als die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Dienstleistung, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse bezieht (ISO 9004-2/8402). Um die Qualität von Prüfungsprozessen, Services und Prüfungen bestimmen zu können, müssen also Erwartungswerte festgelegt werden. Welches Maß an Qualität dabei erreicht werden kann, steht immer auch im Wechselspiel zu den verfügbaren bzw. investierten Ressourcen.¹ Dies gilt gleichermaßen für Strukturen (z. B. Investition in ausfallsichere Hardware), Prozesse (z. B. aufgewendete Supportstunden) als auch Inhalte (z. B. Investition von Arbeitszeit in Konzeption, Erstellung und Qualitätssicherung von Prüfungsfragen).

Bei der Einführung von E-Klausuren an einer Universität kommt dem Risikomanagement eine besondere Bedeutung zu. Zakrzewski und Steven (2000) sowie Ruedel, Schiefner, Noetzli und Schiedt (2007) beschreiben die zyklische Abfolge von Planung, Risikoanalyse, Prüfungsdesign, Implementierung und Evaluation. Auf Grundlage der Evaluationsergebnisse werden die Prüfungsprozesse und das Design angepasst oder erweitert. Bei der Risikoanalyse werden zunächst Problemfelder definiert, diesen Feldern zugeordnete Risiken identifiziert, die Risiken entsprechend dem Produkt aus Schadensausmaß und Eintrittswahrscheinlichkeit klassifiziert bzw. priorisiert und schließlich Maßnahmen zur Risikovermeidung oder Minimierung getroffen. Als Problemfelder bei E-Klausuren lassen sich rechtliche, technische, organisatorische und didaktische Risiken unterscheiden. Abbildung 1 zeigt ein Klassifizierungsschema für die Bereiche *akzeptabel*, *ALARP* (as low as reasonably practicable) und *inakzeptabel*.

¹ Vgl. Magisches Dreieck des Projektmanagements. Abgerufen unter <http://de.wikipedia.org/wiki/Projektmanagement#Stakeholdererwartungen> [28.04.2015]

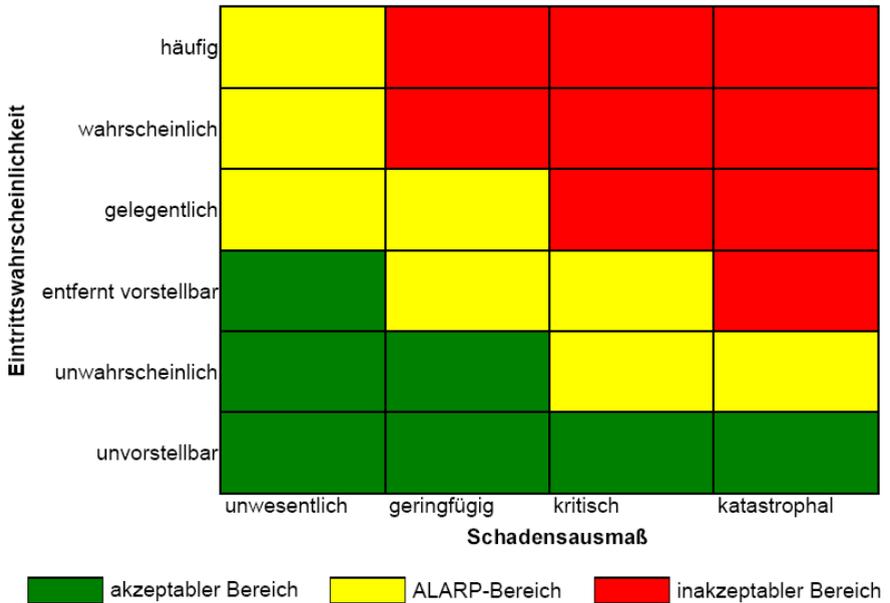


Abbildung 1: Risikoglyph (Quelle: Wikipedia. Abgerufen unter <http://de.wikipedia.org/wiki/Datei:Risikoglyph.png>, verändert [28.04.2015])

Ziel ist es, durch geeignete Maßnahmen die Eintrittswahrscheinlichkeit aller Risiken soweit zu reduzieren, dass sie zumindest in den ALARP-Bereich fallen. Die folgenden Kapitel zeigen beispielhaft für die Problemfelder Informationssicherheit, Organisation und Didaktik einige der im E-Assessment-Service des ZMML etablierten Maßnahmen.

Risiken und Maßnahmen

Informationssicherheit

An die Verfügbarkeit des Prüfungssystems sowie an Sicherheit, Integrität und Schutz der Daten werden bei E-Klausuren sehr hohe Anforderungen gestellt. Mögliche Risiken sind u. a.:

- Prüfungsausfall, -abbruch oder -unterbrechung, z. B. durch Störungen des Netzwerks oder der Stromversorgung
- Datenverluste, z. B. durch Systemausfälle, Software- oder Bedienungsfehler

- Ausfall einzelner Prüfungsstationen, z. B. durch Hardwaredefekte
- Performanceprobleme während der Prüfung, z. B. durch eine fehlerhafte Systemkonfiguration
- Fehlbewertungen durch Softwarefehler
- Verletzung des Datenschutzes, z. B. durch unberechtigte Systemzugriffe

Ohne auf Details eingehen zu können sei auf entsprechende Standards, wie sie z. B. in den IT-Grundschutz-Katalogen (2014) und den IT-Grundschutz-Standards (2008) beschrieben sind, hingewiesen. Wichtige Maßnahmen zur Risikominimierung bei E-Klausuren sind u. a.:

- Hohe Verfügbarkeit der IT-Systeme durch redundante Hardware (Server, Switches etc.) und automatische Übernahme der betroffenen Funktionen bei einem Ausfall
- Geschlossene Netzwerk- und Serverinfrastruktur mit zeitnahe Zugriff der Mitarbeiter auf alle Komponenten und deren Management
- Unterbrechungsfreie Stromversorgung (USV-Anlage) für ein geordnetes Herunterfahren der Server ohne Datenverlust
- Automatisiertes Backup
- Keine lokale Speicherung von Prüfungsdaten auf den Arbeitsplatzrechnern
- Möglichkeit zur direkten Fortsetzung abgebrochener Prüfungen an beliebigen anderen Arbeitsplätzen
- Funktions- und Performanztests nach jeder technischen Änderung und nach relevanten Updates aller betroffenen Systeme (Treiber, Betriebssystem, Datenbanksystem, Webserver, Prüfungssystem etc.)
- Monitoring und regelmäßige Systemwartung
- Zugangsbeschränkungen, z. B. durch Firewall, VPN-Zugänge, rollenbasierte Zugriffsrechte, zeitlich limitierte Accounts etc.

Mit den zuständigen System- und Netzwerkadministratoren sollten verbindliche Vereinbarungen zu Reaktionszeiten und Servicelevel geschlossen werden. Sofern keine eigenen Entwickler für das Prüfungssystem zur Verfügung stehen, ist eine hohe Servicequalität des Systemanbieters (z. B. Hotline, schnelle Reaktion auf Fehlermeldungen) essentiell. An der Universität Bremen ist dies durch die LPLUS GmbH in hohem Maße gegeben.

Zu Beginn des E-Assessment-Services des ZMML wurden die Prüfungen in PC-Pools mit 10 - 24 Plätzen durchgeführt. Diese PC-Pools waren überwiegend dezentral verantwortet, außerhalb der Prüfungszeiten für Studierende frei zugänglich und

mit heterogener Hard- und Software ausgestattet. Hieraus ergaben sich zahlreiche organisatorische und technische Herausforderungen, u. a. eine verteilte Datenhaltung auf verschiedenen Servern, unterschiedliche Prozeduren für Anmeldung und Start der Prüfungsumgebung, keine Möglichkeit, eigene Software wie einen Secure-Browser zu installieren etc. Technische Probleme betrafen insbesondere nicht mehr funktionierende Start-Skripte (z. B. durch den Eingriff von Sheriff-Systemen) sowie verzögerte Prüfungsstarts und kurzzeitige Prüfungsunterbrechungen durch instabile Netzwerkverbindungen (defekte Switches, zeitweiser Ausfall der Breitbandverbindung zum externen Serverstandort). Obwohl es auch in dieser Pionierphase nicht zu Datenverlusten oder Prüfungsausfällen kam, war der personelle, technische und organisatorische Aufwand für die bis zu 7 parallel betriebenen, auf vier Gebäude verteilten Prüfungsräume sowie der geringe Sicherheitslevel für ein dauerhaftes Serviceangebot nicht vertretbar.

Mit Eröffnung des Testcenters im Dezember 2007 konnten alle oben genannten Maßnahmen umgesetzt und die technischen Risiken damit weitestgehend minimiert werden; Prüfungsausfälle, verspätete Starts, technisch bedingte Datenverluste oder gravierende Performanceprobleme traten bisher nicht auf. Mit einem eigenen Warte- und Anmeldebereich, dem 300m² großen Prüfungsraum für die 120 Arbeitsplätze sowie direkt angrenzenden Funktionsräumen bietet das Testcenter alle räumlichen Voraussetzungen für verschiedene Prüfungsszenarien (Bücking & Schwedes, 2007).

Prüfungsprozesse

Die Hauptgeschäftsprozesse bei E-Klausuren lassen sich gliedern in:

- Beratung und Qualifizierung zu Organisation, Software und Prüfungsdidaktik
- Entwicklung und Qualitätssicherung der Fragenkataloge
- Technische Vor- und Nachbereitung der Prüfungen
- Prüfungsdurchführung
- Nachbewertung
- Auswertung und Archivierung der Ergebnisse
- Klausureinsicht

Weitere Prozesse betreffen Anmeldeorganisation, Terminplanung, Organisation der Aufsichten sowie Ergebnisverwaltung (Notenberechnung, Notenbekanntgabe, Archivierung), die sich jedoch nur in Details von denen bei schriftlichen Klausuren unterscheiden.

Jeder dieser Bereiche und die darin enthaltenen Subgeschäftsprozesse haben eigene Risiken und dementsprechend ein Set von Maßnahmen zur deren Vermeidung oder Minimierung. In diesem Beitrag kann aus Platzgründen lediglich auf die Erstellung und Qualitätssicherung der Fragenkataloge detaillierter eingegangen werden (siehe unten bei Prüfungsdesign). Gegenstand dieses Kapitels ist die generelle Ablaufplanung und -kontrolle.

Kennzeichnend für E-Klausuren ist die Komplexität der Interaktionen zwischen Studierenden, Prüfungsverantwortlichen, Fragenautoren, Korrektoren, Aufsichten, verschiedenen Verwaltungseinheiten, IT-Service-Personal und E-Learning-Service. Um die Organisation zahlreicher parallel zu organisierender Prüfungen (aktuell ca. 60 pro Semester) in einem mehrköpfigen Team zu unterstützen, wurde im ZMML das webbasierte Tool eXamControl entwickelt. Dieses bietet u. a. einen passwortgeschützten, verschlüsselten Dateiaustausch, vielfältige Funktionen zur Terminverwaltung und -planung, die Erfassung und statistische Auswertung von Prüfungskennzahlen, eine Nutzerverwaltung sowie die Festlegung und Kontrolle der wichtigsten Aktionen und Deadlines für verschiedene Termintypen. Auf letzteres soll hier näher eingegangen werden.

Während z. B. eine Klausureinsicht lediglich die Aktivierung und anschließende Deaktivierung des Einsichtsmodus für die betreffende Prüfung im Prüfungssystem erfordert, sind für die Vor- und Nachbereitung z. B. eines E-Klausurtermins zahlreiche Deadlines festzulegen und zu kontrollieren. Abbildung 2 zeigt die Planungsansicht für einen gewissen Zeitraum, Abbildung 3 die Aktionsliste für den Termintyp E-Klausur.

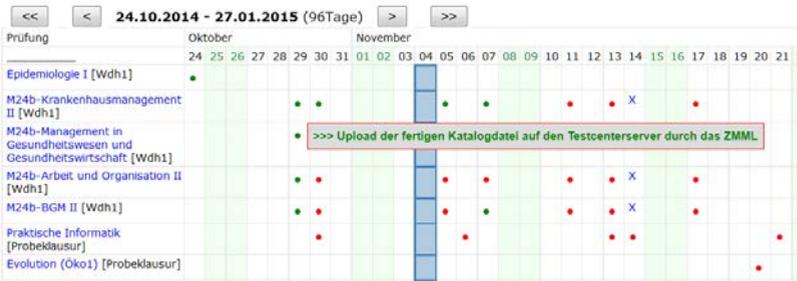


Abbildung 2: Ausschnitt der Planungsansicht in eXamControl zum Status der E-Klausuren eines bestimmten Zeitabschnitts (grün = erledigt, rot = offen)

Prüfungsinfo

Katalog Editor: Unbekannt ZMML Dozent

Nachbewertung:

Aktioneninfo

Terminplan schicken	↔	Mi	01.10.2014	↔	<input checked="" type="checkbox"/>
●					
Fertigstellung der Klausurvorlage + Anlagen, Übermittlung an das ZMML über https://www.zmml.uni-bremen.de/tcfiletool	↔	Mi	15.10.2014	↔	<input checked="" type="checkbox"/>
●					
Fertigstellung des Kataloges, Übermittlung an das ZMML über https://www.zmml.uni-bremen.de/tcfiletool	↔	Mi	22.10.2014	↔	<input checked="" type="checkbox"/>
●					
Upload der fertigen Katalogdatei auf den Testcenterserver durch das ZMML	↔	Mi	29.10.2014	↔	<input checked="" type="checkbox"/>
●					
Einrichten/Aktivieren der VPN-Verbindung zum Testcenter	↔	Do	30.10.2014	↔	<input checked="" type="checkbox"/>
●					
Abschluss Klausurtest (Inhalt, Funktion, Bewertung, Layout)	↔	Mi	05.11.2014	↔	<input checked="" type="checkbox"/>
●					
Abschluss Katalogüberarbeitung + Upload Endversion (ZMML)	↔	Fr	07.11.2014	↔	<input checked="" type="checkbox"/>
●					
Kontrolle der Prüfungseinstellungen (Zeit, Reihenfolge, Auswahl, ...)	↔	Di	11.11.2014	↔	<input type="checkbox"/>
●					
Transfer der TN-Liste an das ZMML / Registrierung im Testcenter	↔	Di	11.11.2014	↔	<input type="checkbox"/>
●					
Freigabe der Prüfung	↔	Do	13.11.2014	↔	<input type="checkbox"/>
●					
Prüfungstermin:		Fr	14.11.2014		
●					
Export + Transfer der Prüfungsergebnisse und Protokolle	↔	Mo	17.11.2014	↔	<input type="checkbox"/>
●					
Katalog archivieren und hochladen	↔	Mo	17.11.2014	↔	<input type="checkbox"/>

▼

Abbildung 3: Aktionsliste und Deadlines für einen E-Klausurtermin in eXamControl

Jedes Teammitglied kann den Status der Aktionen von offen auf erledigt setzen, die Prüfungsverantwortlichen selber können, bei den aktuell genutzten Rechteeinstellungen, den Status ihrer Prüfungen zwar nicht ändern, aber einsehen.

Jede in eXamControl angelegte Prüfung ist mit einem korrespondierenden Ticket in RT - Request Tracker (<https://bestpractical.com>), einem open source – Anfragesystem verknüpft. Diese Tickets können direkt aus eXamControl heraus generiert werden, so dass alle registrierten Prüfungsverantwortlichen als Klienten eingetragen und in die Benachrichtigungsfunktionen von RT einbezogen werden. Da RT

auch E-Mail-basiert gesteuert werden kann, lassen sich z. B. die in eXamControl generierten Ablaufpläne einfach an den RT-Verteiler aller betroffenen Lehrenden und Servicemitarbeiter versenden. Der gesamte E-Mail-Verkehr zu einer Prüfung läuft dann ausschließlich über das betreffende Ticket. Außerhalb liegende Kommunikationsprozesse, z. B. Kommentare zu telefonischen Absprachen, versehentlich neu per E-Mail an die Supportadresse erzeugte Tickets oder an andere E-Mail-Adressen gesendete Anfragen zu einer Prüfung, lassen sich einfach in das korrekte Ticket integrieren. Die so gebündelten Informationsflüsse erlauben allen Servicemitarbeiter / Innen jederzeit Einblick in den aktuellen Status der Kommunikation und sind Teil der Prüfungsdokumentation.

Prüfungsdesign

Mögliche Risiken für die Ergebnisqualität von E-Klausuren im Sinne der klassischen Hauptgütefaktoren Validität, Objektivität und Reliabilität sind u. a.:

- Fokussierung der erfassten Kompetenzen auf Informationserinnerung und ein damit verbundener, negativer Lernlenkungseffekt
- fehlende Berücksichtigung und Bewertung von Lösungswegen
- zu geringe Investition von Arbeitszeit in Fragenerstellung und Qualitätssicherung
- mangelnde Berücksichtigung von im Sinne des Constructive Alignments² wichtiger Kompetenzen durch den Verzicht auf offene Fragenformate bzw. durch Vermeidung des damit verbundenen Nachbewertungsaufwands
- fehlerhafte Fragen hinsichtlich Funktion, Fragestellung oder Bewertung.

Auch wenn von Seiten der Servicestelle kein direkter Einfluss auf Prüfungsinhalte genommen werden kann und sollte, setzen die vorgelagerten und begleitenden Beratungs- und Supportprozesse wichtige Impulse für eine Qualitätssteigerung. Wichtige Elemente dabei sind die prüfungsdidaktische Beratung, die Bereitstellung von Musterkatalogen, Fragenvorlagen und Checklisten und die direkte Einbindung des ZMML in den Qualitätssicherungszyklus der Fragenerstellung. Auch das Prüfungssystem bzw. die darin bereitgestellten Werkzeuge für das Testen und Überarbeiten von Fragen durch Autorenteam spielen eine Rolle.

² Zum Konzept des Constructive Alignments siehe Biggs und Tang (2011)

Beratung

Im initialen Beratungsgespräch werden neben Fragen der Organisation und Technik insbesondere die prüfungsdidaktischen Möglichkeiten und Grenzen, nach Möglichkeit auf Grundlage bereits vorhandener Fragenvorlagen, besprochen. Besonderes Augenmerk wird dabei auf das Primat der zu überprüfenden Kompetenzen gegenüber den technischen Möglichkeiten gelegt.

Fragen, deren Beantwortung z. B. umfangreiche mathematische Herleitungen oder komplexe Zeichnungen erfordert, lassen sich in E-Klausuren nicht effizient abbilden. Hier bieten sich hybride Prüfungen an, in denen diese Fragen auf Papier, alle übrigen wie gewohnt am Computer beantwortet werden. Die Integration der Anwendung fachspezifischer Software in eine E-Klausur erlaubt es, zusätzlich zu Informationserinnerung, -verarbeitung und -erzeugung auch berufspraktische Kompetenzen zu prüfen. Ein typisches Beispiel ist die E-Klausur *Statistik für Wirtschaftswissenschaftler* (Prof. Missong, Fachbereich Wirtschaftswissenschaft), in der der Großteil der Fragen in den typischen Formaten der E-Klausur gestellt wird (Auswahlfragen, Eingabefelder für Berechnungen, Freitext), einige wenige komplexe Herleitungen auf Papier eingereicht werden und für alle Fragen eine Tabellenkalkulation sowie spezielle Statistiktools als Werkzeuge genutzt werden können.

Support-Materialien

Im Beratungsgespräch erhalten die Prüfungsverantwortlichen einen Zugang zu dem System eXamControl, in dem im Dateibereich neben Software, Handbüchern und Kurzanleitungen auch eine umfangreiche Sammlung von Fragenvorlagen und Checklisten bereitgestellt wird. Für Autoren, die das ZMML mit der Digitalisierung der Fragen beauftragen, stehen nach Fragetypen sortierte Formulare bereit, in denen neben obligatorischen Elementen wie Fragestellung, Thema, Bewertung und Antwortoptionen optional Angaben zu Lernziel, Lernzielniveau nach Anderson und Krathwohl (2001) oder nach Metzger und Nüesch (2004), Szenario-Beschreibung, Bedienungs- und Bewertungshinweise gegeben werden können. Die Vorlage erlaubt zudem die Erstellung von Blueprints für die Prüfungszusammenstellung mit einem Vorschlag für eine Quotierung nach Lernzielniveau.

Autoren, die nach entsprechender Schulung ihre Fragen im LPLUS-Editor digitalisieren, finden nach der Installation einen integrierten Vorlagenkatalog des ZMML mit Musterfragen vor, die direkt in den eigenen Katalog kopiert und angepasst werden können. Platzhalter für alle Elemente inkl. Szenario bzw. Handlungssituation, Arbeitsanweisung und Bewertungshinweis sorgen dafür, dass diese bei der Erstellung eher berücksichtigt werden. Gleichzeitig unterstützt die Vorlage ein gutes Fragendesign hinsichtlich Layout, Funktion und korrekte Bewertheinstel-

lungen und weist indirekt auf die Vielfalt möglicher Fragetypen und Konstruktionen hin.

Vorlagenkatalog und Fragentemplate sind unter <http://www.eassessment.uni-bremen.de/supportunterlagen.php> frei zugänglich.

Qualitätssicherungszyklus für die Fragenkataloge

Eine Besonderheit von E-Klausuren gegenüber schriftlichen Prüfungen liegt in der Vorverlagerung der Bewertung bei Fragen im Antwort-Wahl-Verfahren. Die Erstellung der Fragenkataloge und des Prüfungsdesigns muss hier deutlich vor dem eigentlichen Prüfungstermin erfolgen. Um formale, technische und inhaltliche Fehler und das daraus resultierende Risiko eines hohen Nachbewertungsaufwands zu vermeiden, erhält die vorherige Kontrolle der Fragen deutlich mehr Gewicht. Abbildung 4 zeigt die an der Universität Bremen etablierte Vorgehensweise.

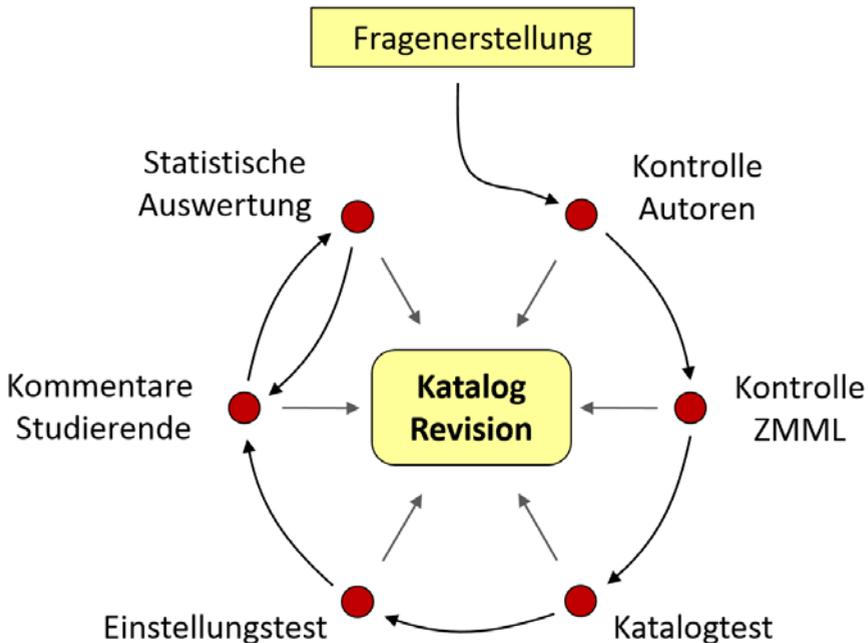


Abbildung 4: Qualitätssicherung der Fragenkataloge

Prüfkriterien für die Fragen (lokale Kontrollen der Offline-Version und Online-Katalogtest) sind u. a. Layout und Funktion, Inhaltsvalidität, ungewollte Lösungs-

hinweise, thematische Zuordnung und Bewertungseinstellungen. Der Test der Prüfungseinstellungen fragt u. a. nach Prüfungszeit und Auswahl, Reihenfolge und Bündelung der Fragen sowie Optionen wie Hinweistext bei Prüfungsstart, Ergebnisanzeige und Taschenrechneranzeige. Die Kommentarfunktion erlaubt es den Studierenden, während der Prüfung auf Probleme wie Unklarheiten bei Fragestellung, Bewertung und Bedienung oder fehlende Antwortoptionen hinzuweisen. Diese wertvollen Kommentare dienen zusammen mit den Ergebnissen der Prüfungs- und Fragenstatistik (Trennschärfe, Schwierigkeitsgrad, Antworthäufigkeiten) der Verbesserung des Fragenkataloges oder weisen auf eine ggf. notwendige Neubewertung hin. Bei einer Zufallsauswahl von Fragen kann über diese Kennzahlen die rechtlich geforderte Gleichwertigkeit der aus einem Pool gezogenen Fragen überprüft werden. Mit jeder weiteren Prüfung bzw. Verwendung der Fragen bietet sich so die Möglichkeit für erneute Anpassungen des Fragenkataloges in einem kontinuierlichen Verbesserungsprozess.

Qualitätssicherungsfunktionen des Prüfungssystems

Das LPLUS-Teststudio 2.0 (LTS, s. www.lplus.de) erlaubt einer beliebigen Anzahl von Überarbeitern einfache Bewertungsfunktionen, Annotationen, E-Mail-Benachrichtigungen und die abgestufte Freigabe von Fragen und Fragenauswahlen zu vorab festgelegten Deadlines. Abbildung 5 zeigt einen Screenshot des Vorlagen-tests im LTS.

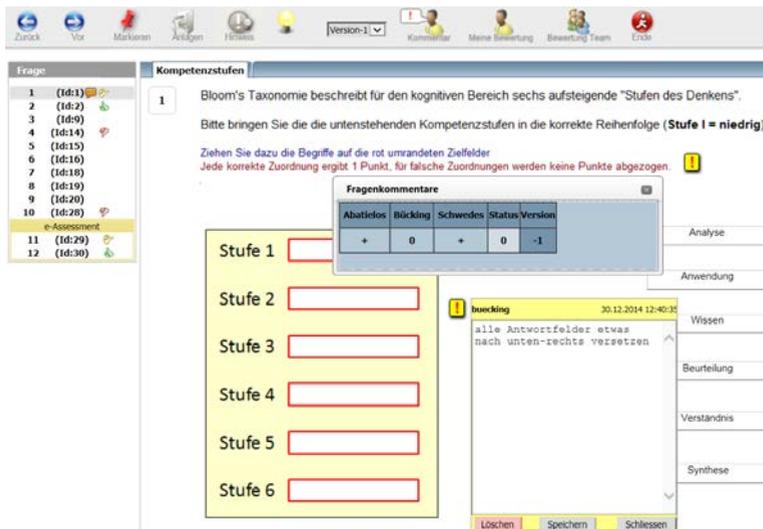


Abbildung 5.: Vorlagentest im LPLUS-Teststudio 2.0 (Screenshot)

Die auf Basis der hier hinterlegten Änderungskommentare notwendigen Korrekturen an Fragen und Einstellungen werden in diesem Stadium der Qualitätskontrolle nur noch durch das ZMML vorgenommen. Erst nach dem Prüfungstermin erhalten die Lehrenden zusammen mit den Ergebnisdaten auch die aktuelle Katalogdatei für die weitere Aktualisierung und Ausbau.

Fazit

Im Laufe von 10 Jahren E-Assessment an der Universität Bremen traten die technischen und organisatorischen Probleme mit zunehmender Routine immer mehr in den Hintergrund. Lediglich bei Innovationen wie neuen Software-integrierten Prüfungssettings oder dem aktuell stattfindenden Austausch aller Server wird erneut in den Risiko-Management-Zyklus eingetreten. Das Testcenter als infrastrukturelle Basis für den Dienst E-Assessment hat sich an der Universität Bremen durchgehend bewährt. Das Thema Prüfungsdidaktik hat demgegenüber immer mehr Gewicht bekommen. Die Frage, wie auch mit automatisierten Bewertungsverfahren kompetenzorientierte formative und summative Prüfungen gestaltet werden können ist hochaktuell. Die vorhandenen Beratungs- und Schulungsangebote sollten weiter ausgebaut und z. B. in hochschuldidaktische Zertifikatsprogramme eingebunden werden. Der sich daraus ergebene zusätzliche Ressourcenbedarf lässt sich ggf. durch standortübergreifende Kooperationen reduzieren. Zusammenfassend lässt sich festhalten, dass zu den im magischen Dreieck genannten Faktoren Qualität, Ressourcen und Termin ein viertes Element, nämlich das Erfahrungswissen, hinzutritt. Dieses hat sowohl auf Seiten der Nutzer als auch der Serviceanbieter einen entscheidenden Einfluss auf die Qualität aller Prüfungsprozesse und reduziert zunehmend den hohen initialen Ressourceneinsatz.

Referenzen

- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university: What the student does* (4. Aufl.). Maidenhead: McGraw Hill Education and Open University Press.
- Bundesamt für Sicherheit in der Informationstechnik (2008). *Informationssicherheit und IT-Grundschutz: BSI-Standards 100-1, 100-2 und 100-3*. Köln: Bundesanzeiger-Verlag.
- Bundesamt für Sicherheit in der Informationstechnik (2014). *IT-Grundschutz-Kataloge: Standardwerk zur IT-Sicherheit*. Köln: Bundesanzeiger-Verlag.

Bücking, J. & Schwedes, K. (2010). E-Assessment im Testcenter der Universität Bremen. In C. Ruedel & S. Mandel (Hrsg.), *E-Assessment: Einsatzszenarien und Erfahrungen an Hochschulen* (Medien in der Wissenschaft, Bd. 56, S. 47-62). Münster: Waxmann Verlag.

Metzger, C. & Nüesch, C. (Hrsg.). (2004). *Fair prüfen: Ein Qualitätsleitfaden für Prüfende an Hochschulen* (Hochschuldidaktische Schriften, Bd. 6). St. Gallen: Institut für Wirtschaftspädagogik der Universität St. Gallen.

Ruedel, C., Schiefner, M., Noetzi, C. & Seiler Schiedt, E. (2007). Risikomanagement für E-Assessment. In M. Merkt, K. Mayrberger, R. Schulmeister, A. Sommer & I. van den Berg (Hrsg.), *Studieren neu erfinden – Hochschule neu denken* (Medien in der Wissenschaft, Bd. 44, S. 180-190). Münster: Waxmann.

Zakrzewski, S. & Steven, C. (2000). A Model for computer-based assessment: The Catherine Wheel Principle. *Assessment & Evaluation in Higher Education*, 25(2), 201-215.

Vita

Dr. Jens Bücking ist seit 2003 Mitarbeiter des Zentrums für Multimedia in der Lehre (ZMML) der Universität Bremen. Sein Aufgabenbereich umfasst das Projekt- und Servicemanagement mit dem Schwerpunkt E-Assessment. Seit 2004 verantwortet er die Durchführung elektronischer Klausuren. Vor seinem Wechsel zum ZMML war er nach Studium und Promotion in der Biologie an der Universität Bremen 12 Jahre in Forschung und Lehre tätig.

Susanne Bergann
Freie Universität Berlin

Wie beurteilen Studierende computergestützte Prüfungen? – Erste Ergebnisse der Evaluation der E-Examinations an der Freien Universität Berlin

Um die Qualitätssicherung und -entwicklung durch die Bereitstellung empirischer Daten zu unterstützen, werden die seit dem Wintersemester 2012 / 2013 im E-Examination Center der Freien Universität Berlin durchgeführten computergestützten Prüfungen durch regelmäßige Evaluationen flankiert. In der ersten Phase der Evaluation stand dabei die Frage im Vordergrund, wie die Studierenden das Konzept der Computerisierung des Prüfungswesens generell bewerten. Des Weiteren wurde erfasst, wie zufrieden die Studierenden mit den Rahmenbedingungen der Prüfung waren und inwieweit etwaige technische Probleme während der Prüfung aufgetreten sind. Zusätzlich wurden soziodemografische Merkmale sowie computer- und prüfungsrelevante Merkmale erhoben (z. B. computerrelevante Vorkenntnisse, Prüfungsangst). Im ersten Evaluationsdurchgang (WS 2013 / 2014) haben insgesamt $N = 1049$ Studierende aller Fachbereiche der Freien Universität Berlin an der Befragung teilgenommen. Insgesamt weisen die Ergebnisse darauf hin, dass die Befragten eine moderat positive Einstellung gegenüber computergestützten Prüfungen haben und nur etwas mehr als die Hälfte der Befragten angab, klassische Prüfungen gegenüber E-Examinations zu präferieren. Weiterführende Analysen zeigen weiterhin, dass insbesondere computerbezogene Merkmale (computerrelevante Vorkenntnisse, computerbezogene Selbstwirksamkeitserwartung) sowie die Zufriedenheit der Studierenden mit den Rahmenbedingungen der Prüfung mit der Bewertung der E-Examinations in Zusammenhang stehen.

Prüfungssoftware

Vergleich der gängigen Prüfungssoftwarelösungen

Zusammenfassung

In diesem Beitrag werden als Prüfungssoftwarelösungen die Open Source E-Learning-Systeme ILIAS und Moodle und das kommerzielle Online-Prüfungssystem LPLUS miteinander verglichen. Es geht übersichtsartig um Fragetypen, Bewertungseinstellungen und das Prüfungsmanagement.

Einleitung

Bewertungsansatz

Die Open Source E-Learning-Systeme ILIAS und Moodle und das kommerzielle Online-Prüfungssystem LPLUS werden gegenübergestellt. Es geht übersichtsartig um Fragetypen, Bewertungseinstellungen und das Prüfungsmanagement.

Alle Systeme kennen nicht nur Single- und Multiple-Choice-Fragen, sondern auch Zuordnungen, Anordnungen, Lückentext- und Hotspotfragen, Numerische und Berechnungsfragen sowie Freitextaufgaben.

Spezifische Fragetypen, die nicht alle betrachteten Systeme kennen, sind darüber hinaus z. B.: Fehler / Worte markieren, zusammenhängende Aufgaben (Cluster) etc.

Bei den Bewertungseinstellungen liegen die größten Unterschiede und daraus folgend die eigentliche individuelle Entscheidung für oder gegen ein Prüfungs- bzw. Test-System.

Meinem Vergleich liegen folgende Softwareversionen zu Grunde:

- ILIAS: 4.3.9
- Moodle: 2.6.x
- LPLUS: TM Editor 2013.10.0, TS-Editor Beta

Single- / Multiple-Choice

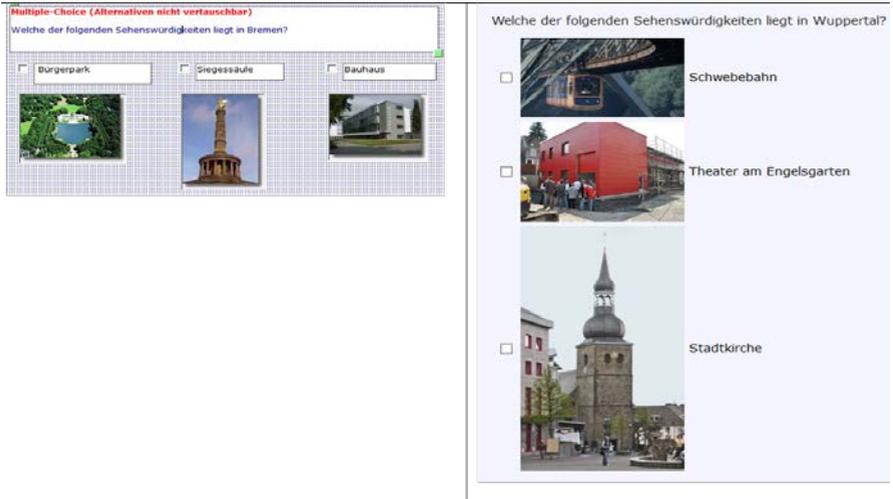


Abbildung 1: Beispiel für eine Multiple-Choice-Frage in LPLUS und ILIAS

Bei Multiple-Choice-Fragen (MC) können mehrere, alle oder keine der Antworten richtig sein.

Multiple-Choice-Frage mit Auswahlpaaren

Alle verglichenen Systeme bieten eine Variante von Multiple-Choice-Fragen, bei der jeweils zwei Antworten als Auswahlpaar gewertet werden, d. h. es ist nur möglich, pro Paar a oder b anzukreuzen (Kprim).

Ja/Nein mit SingleChoice

- Aussagen zu Personalwesen und Organisation
 Personal- und Organisationsstrukturen setzen das Unternehmensprogramm um.
- Wahr
 Falsch



Frage 1

Bisher nicht
beantwortet

Erreichbare Punkte:
1,00

Beurteilen Sie diese Aussage:

XX

Eine auswählen:

- Wahr
 Falsch

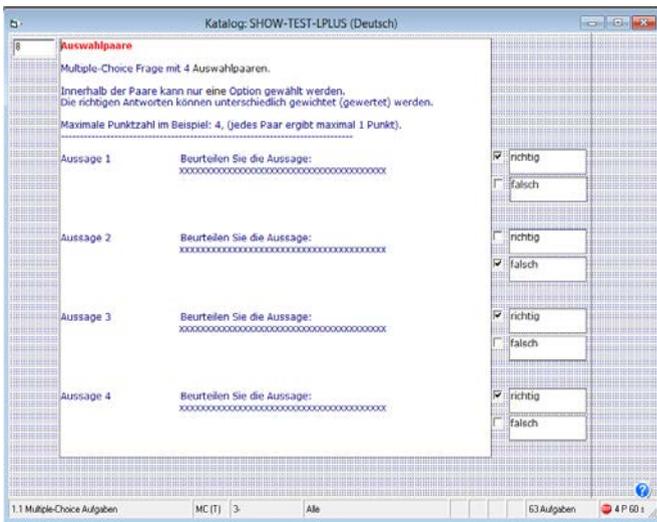


Abbildung 2: Beispiel einer Kprim-Aufgabe in ILIAS, Moodle und LPLUS

Bewertungseinstellungen bei MC-Fragen

Bei Mehrfachauswahl-Fragen vergibt ILIAS sowohl für die Auswahl als auch für die Nicht-Auswahl einer Antwortmöglichkeit Punkte.

Bewertungssystem

- Teillösungen bewerten
- Nur vollständig richtige Lösungen bewerten

Wenn Teillösungen bewertet werden, werden die definierten Punkte abhängig von den Antworten der Teilnehmer aufsummiert. Wenn nur vollständige Lösungen bewertet werden, müssen Teilnehmer die maximale Punktezahl für eine Frage erzielen, anderenfalls wird die Frage mit 0 Punkten bewertet.

Multiple Choice-Fragen

- Vergabe von 0 Punkten, wenn keine Antwort ausgewählt wurde
- Vergabe der Punktezahl, auch wenn keine Antwort ausgewählt wurde

Abhängig von der Punktevergabe bei Multiple Choice Fragen kann es passieren, dass für Fragen, bei denen keine Antwort ausgewählt wurde, Punkte vergeben werden. Dies wird durch die Standardeinstellung verhindert, indem in diesem Fall die Frage als nicht beantwortet gewertet wird und die Punkte automatisch auf Null gesetzt werden.

Negative Punkte abschneiden

- Negative Punkte für alle Fragen bei Null abschneiden
- Negative Punkte für die Gesamtsumme aller Fragenergebnisse bei Null abschneiden

Negative Punkte können für einzelne Fragen oder aber für die Summe aller Fragenergebnisse auf 0 zurückgesetzt werden. Bei der Summe aller Fragenergebnisse fallen negative Punkte stärker ins Gewicht.

Abbildung 3: Bewertungseinstellungen in ILIAS bei MC-Fragen

Bewertungen in Moodle festlegen:

- Bewertung für einzelne Fragen: Die Punktzahl ist in das Textfeld „Erreichbare Punkte“ unter der Frage einzutragen:

Aktuelle Kategorie MC-Beispiele (7) Diese Kategorie benutzen

In der Kategorie sichern MC-Beispiele (7)

Fragetitel* Welcher Dichter schrieb folgendes Gedicht?

Fragetext*

Absatz **B** *I*

Ein Nagel saß in einem Stück Holz.
 Der war auf seine Gattin sehr stolz.
 Die trug eine goldene Haube
 Und war eine Messingschraube.

Format: p

Erreichbare Punkte*

Abbildung 4: Bewertungseinstellungen in Moodle bei MC-Fragen für die einzelne Frage

- Bewertung für den Test: Legen Sie die Bewertung für den gesamten Test fest: Tragen Sie die Gesamtpunktzahl im Textfeld „Beste Bewertung“ oberhalb der Fragenliste ein.

Test bearbeiten: Beispiel Test Test Test

Die Basisideen zur Testerstellung
 Summe der Bewertungen: 10,00 | Fragen: 6 | Aktuell läuft dieser Test
 Beste Bewertung: 10,00

Seite 1

1 1. Zigarette am Morgen Wann nach dem Aufwachen rauchen Sie Ihre e... Erreichbare Punkte: 3

Multiple-Choice

Seite hier hinzufügen

Seite 2

2 2. Rauchverbot einhalten? Finden Sie es schwierig, an Orten, wo das Rau... Erreichbare Punkte: 1

Multiple-Choice

Seite hier hinzufügen

Abbildung 5: Bewertungseinstellungen in Moodle bei MC-Fragen für den Test

Im LPLUS-System sind die möglichen und zu berücksichtigenden Einstellungen recht komplex:

Mehrere richtige Antworten ohne Teilbewertung, „Alles oder Nichts“

Wenn mehrere Antworten korrekt sind und keine Teilbewertung zugelassen ist, gibt es keine Punkte, wenn nicht alle korrekten Antworten gefunden wurden. Sobald eine falsche oder nicht alle korrekten Antworten angekreuzt werden, gibt es für die Aufgabe 0 Punkte.

Mehrere richtige Antworten mit Teilbewertung, Option 1

Mit der Option 1 (Ankreuzen falscher Antworten gibt grundsätzlich 0 Punkte (Keine Verrechnung)) sind auch Teilpunkte möglich. Jede korrekte Antwort sollte dabei mit einer Punktzahl versehen sein. Es ist somit möglich, nur eine von zwei Lösungen anzuhaken und die Teilpunkte der jeweiligen Antwort zu bekommen.

Mehrere richtige Antworten mit Teilbewertung, Option 2, „Verrechnung“

Bei gewählter Option 2 (Teilpunkte falscher Antworten werden abgezogen) sollten alle Antwortalternativen mit einer Punktzahl versehen sein (falsche Alternativen dürfen 0 Punkte haben – dann gibt es keine Abzüge). In der Regel wird für jede falsche Antwort aber eine Punktzahl vergeben. Sollte nun eine falsche Antwort ausgewählt werden, bekommt man einen Minuspunkt, der ggf. mit bereits erreichten Punkten verrechnet wird (bis zum Minimum von 0 Punkten insgesamt).

Malusregelung ist hierbei zu beachten!

Mehrere richtige Antworten mit Teilbewertung, Option 3, „Verrechnung“

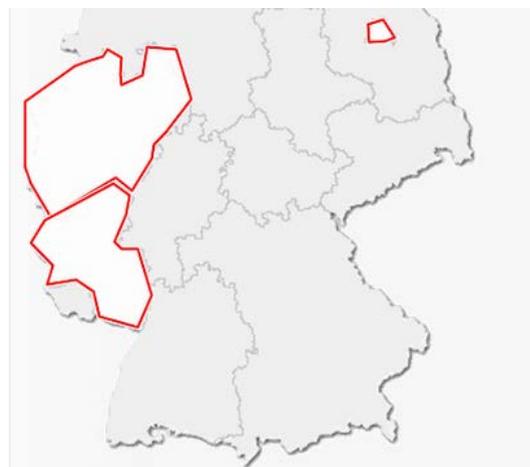
Bei der Option 3 der Teilbewertung (0 Punkte, wenn mehr Antworten angekreuzt werden als Richtige vorhanden sind) wird geprüft, ob mehr Antworten angekreuzt werden, als Richtige vorhanden sind. Sind weniger oder genau die richtige Anzahl an Antworten angehakt, gibt das Programm für die korrekten Antworten Punkte.

Abbildung 6: Bewertungseinstellungen in LPLUS bei MC-Fragen

Hotspot / Imagemap

Hierbei laden Sie ein Bild von Ihrem lokalen Rechner auf die ILIAS-Installation und definieren darin klickaktive Felder („Regionen“), unter denen die Testpersonen durch Anklicken im Bild die richtige finden müssen.

Bei ILIAS werden für jeden richtigen Hotspot die erreichbaren Punkte eingetragen:



Bestehende Datei löschen

Hinweis	Punkte	Form
Alternative 1	1	POLY 47,190, 58,186, 71,182, 75,176, 8
Alternative 2	0	POLY 299,155, 312,151, 322,166, 314,1
Alternative 3	0	POLY 94,299, 91,315, 81,338, 87,344, 1

Abbildung 7: Hotspot / Imagemap-Frage in ILIAS

Moodle benötigt ein Plugin:

- Plugin Question types: Drag and drop markers
- qtype_ddmarker
- Markierungen müssen auf der Hintergrundgrafik an der richtigen Stelle platziert werden

Beim TM-Editor von LPLUS existieren vier verschiedene Ausprägungen von Hotspot-Aufgaben.

- *Einfacher Hotspot*: Ein bestimmter Zielbereich wird auf der voreingestellten Hintergrundgrafik markiert, diese Markierung muss zur Lösung der Aufgabe angeklickt werden.
- *Alternative Hotspots*: Aus mehreren Zielbereichen muss ein Bereich angeklickt werden.
- *Gruppenhotspots*: Alle Zielbereiche müssen angeklickt werden.
- *Hotspot-Sequenz*: Die Zielbereiche müssen in der vordefinierten Reihenfolge angeklickt werden.

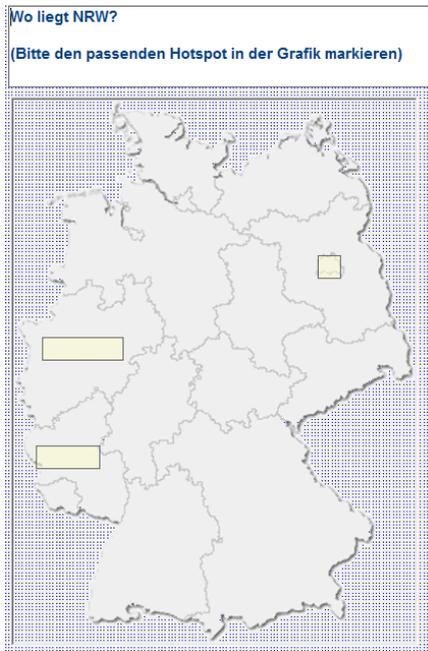


Abbildung 8: Hotspot-Frage im LPLUS TM-Editor

In LPLUS beinhalten die Lösungseinstellungen für jeden Hotspot die gewünschte Punktzahl und die Position (nur wirksam innerhalb einer Sequenz).

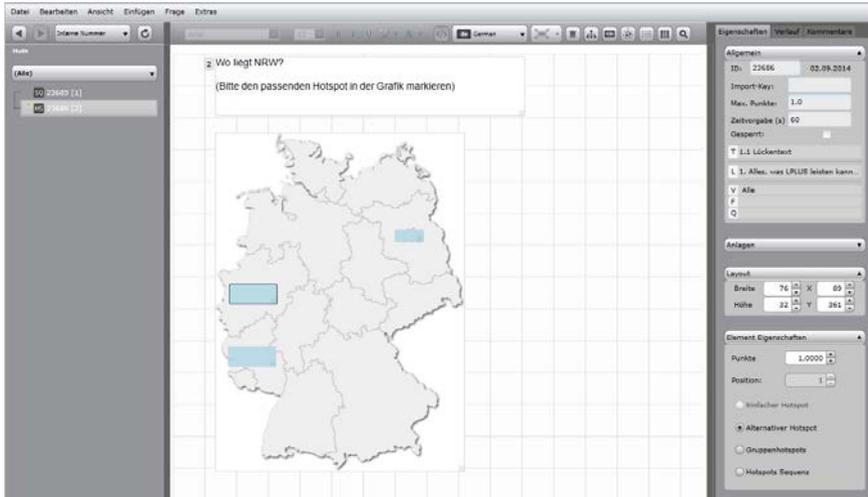


Abbildung 9: Hotspot-Frage im neuen LPLUS TS-Editor

Während ILIAS für die Hotspot-Antwortelemente beliebige Polygone anbietet, können diese bei LPLUS (bisher noch) nur rechtwinklig erzeugt werden, was bei einer Bundesländerkarte wie in den gezeigten Beispielen nicht gut passt.

Lückentextfrage

In ILIAS bestehen diese Fragen aus einer Textpassage (mit Code zum Erzeugen der Lücken), in der die Antworten entweder aus Text-Lücken oder Auswahl-Lücken bestehen.

- Text-Lücken werden für die Testpersonen als leeres Eingabefeld angezeigt. Die Testpersonen müssen hier einen freien Text eintragen, der mit den vorgegebenen Antworten verglichen wird. Bei der Erstellung von Text-Lücken werden alle möglichen Lösungen angegeben, die eine Testperson eintragen darf. Nur wenn die eingegebene Lösung mit einem der eingetragenen Begriffe übereinstimmt, wird eine Textlücke als korrekt beantwortet gewertet.
- Bei Auswahl-Lücken wird den Testpersonen ein Drop-Down-Menü mit vorgegebenen Antworten angezeigt, aus denen die korrekte Antwort ausgewählt werden muss. Bei der Erstellung von Auswahl-Lücken werden die richtige und eine oder mehrere falsche Antworten eingetragen.

Aus dem Leitbild

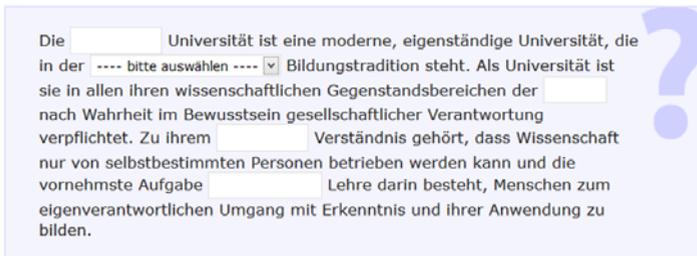


Abbildung 10: Lückentext-Frage in ILIAS

In ILIAS werden richtige Lösungen je nach Lückentyp unterschiedlich angegeben und bewertet.

Lücke 1			
Typ	Textlücke		
Werte *	Antwort-Text	Punkte	Aktionen
	Bergische	1	+ -
Lücke 2			
Typ	Auswahl-Lücke		
Werte *	Antwort-Text	Punkte	Aktionen
	humboldtschen	1	+ -
	humanistischen	0	+ -
Antworten mischen	<input checked="" type="checkbox"/>		

Abbildung 11: Bewertungseinstellung von Lückentext-Fragen in ILIAS

Beispiel in Moodle:

Frage 1
Bisher nicht beantwortet
Erreichbare Punkte: 5,00

Die Universität ist eine moderne, eigenständige Universität, die in der Bildungstradition steht. Als Universität ist sie in allen ihren wissenschaftlichen Gegenstandsbereichen der nach Wahrheit im Bewusstsein gesellschaftlicher Verantwortung verpflichtet. Zu ihrem Verständnis gehört, dass Wissenschaft nur von selbstbestimmten Personen betrieben werden kann und die vornehmste Aufgabe Lehre darin besteht, Menschen zum eigenverantwortlichen Umgang mit Erkenntnis und ihrer Anwendung zu bilden.

Abbildung 12: Lückentext-Frage in Moodle

Eingabe dieses Beispiels:

Abwatz

Die (1 SHORTANSWER =Bergische) Universität ist eine moderne, eigenständige Universität, die in der (1 MULTICHOICE =humboldtischen-humanistischen) Bildungstradition steht. Als Universität ist sie in allen ihren wissenschaftlichen Gegenstandsbereichen der (1 SHORTANSWER =Suche) nach Wahrheit im Bewusstsein gesellschaftlicher Verantwortung verpflichtet. Zu ihrem (1 SHORTANSWER =ethischen) Verständnis gehört, dass Wissenschaft nur von selbstbestimmten Personen betrieben werden kann und die vornehmste Aufgabe (1 SHORTANSWER =akademischer) Lehre darin besteht, Menschen zum eigenverantwortlichen Umgang mit Erkenntnis und ihrer Anwendung zu bilden.

Format p

Abbildung 13: Eingabe einer Lückentext-Frage in Moodle

LPLUS kennt bisher noch keinen eigenen Fragentyp Lückentext, bietet hierfür aber den Typ Sonderfrage an, der mit vielfältigen Möglichkeiten ausgestattet ist.

Katalog: Test-Lückentext (deutsch)

4 Füllen Sie den Lückentext aus:

Was man unter Erziehung versteht, ist die zentrale Frage der Pädagogik.

Als hung bezeichnet n in r Wissenschaft le bewussten d gezielten lungen und nsweisen eines tiv erfahrenen chen.

Abbildung 14: Lückentext-Frage als Sonderfrage-Typ in LPLUS

Der kommende TS-Editor von LPLUS definiert als besondere Form der Sonderfrage jetzt Lückentexte:



Abbildung 15: Lückentext-Frage im TS-Editor von LPLUS

Mit ihm lassen sich auch teilbewertete Lückentextaufgaben anlegen.

Freitext-Frage

ILIAS erlaubt die Festlegung einer maximalen Anzahl von Zeichen:

Ihr Fazit

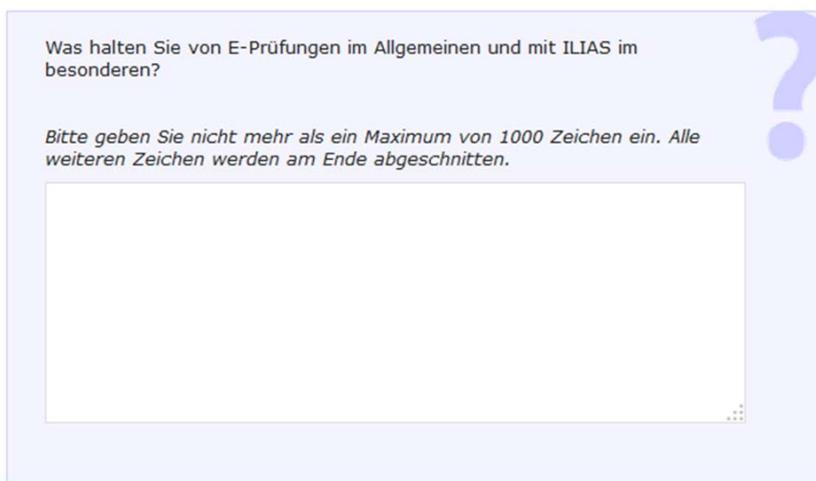


Abbildung 16: Die Frageform Freitext in ILIAS

Die Festlegung einer maximalen Anzahl von Zeichen kennen Moodle und LPLUS nicht.

Wichtig ist insbesondere die Frage der Datenspeicherung bei langen Antworten:



Abbildung 17: Automatisches Speichern in ILIAS

Moodle: Die Aktivität *Test* speichert nur alle fünf Minuten automatisch. Auf der Administrationsebene lässt sich unter “*Plugins* → *Aktivitäten* → *Tests*” die Zeitspanne zur automatischen Sicherung auf eine Minute herabsetzen.

Auch LPLUS speichert automatisch jede Minute im Hintergrund.

Übersicht der Fragetypen

Tabelle 1: Übersicht der Fragetypen in den betrachteten Prüfungssoftwaresystemen

Fragetyp	ILIAS	Moodle	LPLUS
Single-Choice	✓	✓	✓
Multiple-Choice	✓	✓	✓
Multiple-Choice-Frage mit Auswahlpaaren	✓	✓	✓
Zuordnungsfrage	✓	✓	✓
Anordnungsfrage - vertikal	✓		✓
Anordnungsfrage - horizontal	✓		✓
Lückentextfrage	✓	✓	(✓)
Hotspot / Imagemap	✓	(✓)	✓
Fehler / Worte markieren	✓		(✓)
Begriffe benennen	✓		✓
Numerische Antwort	✓	✓	✓
Formel- / Berechnete Fragen	✓	✓	
Freitext	✓	✓	✓
Zusammenhängende Aufgaben (Cluster)			✓

Prüfungsmanagement

In ILIAS stehen unterschiedliche Übersichten zur Verfügung:

- Die Übersicht eines ausgewählten Teilnehmers enthält die Anzahl der Durchläufe, das Datum, die Anzahl der beantworteten Fragen, die erreichte Punktzahl sowie den Prozentsatz der gelösten Aufgaben.

Fragen Info Einstellungen Teilnehmer Lernfortschritt Manuelle Bewertung Statistik Verlauf Metadaten Export

Rechte

Zurück Drucken PDF-Export

Übersicht der Testdurchläufe

Name: Dieter Huth
Datum des Tests: 2014-08-08 16:07:25

Testergebnisse

Bewerteter Durchlauf	Durchlauf	Datum	Beantwortete Fragen	Erreichte Punkte	Prozent gelöst	
⊙	1	08. Aug 2014	13 von 13	27 von 41	65.85%	Aktionen ▾

Herzlichen Glückwunsch! Sie haben den Test bestanden und dabei die Note "bestanden" erzielt.

Abbildung 18: ILLIAS: Übersicht der Testdurchläufe

Moodle:

- Im Bericht *Test-Bewertung* wird für jeden einzelnen Testversuch angezeigt, wie viele Punkte insgesamt bzw. für jede einzelne Frage erreicht wurden, wann der Testversuch begonnen bzw. beendet wurde, wie viel Zeit benötigt und ob der Testversuch abgeschlossen wurde. Ein Histogramm zeigt alle erreichten Bewertungen.

Was in Bericht einbezogen wird

Versuche von Eingetragene mit Testversuchen
 Versuche von In Bearbeitung Überfällig Beendet Nie zugegriffen
 die neu bewertet bzw. dafür markiert wurden

Anzeigegrößen

Beleggröße
 Bewertungen für jede Frage

In diesem Teil ist nur ein Versuch pro Nutzer möglich.

Vorname: ABBACDDEFGHIJKLMNOPQRSTUVWXYZ
 Nachname: ABBACDDEFGHIJKLMNOPQRSTUVWXYZ
 Seite: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 (Weiter)

Belegnr.	Vorname	E-Mail-Adresse	Status	Begonnen am	Beendet	Verbrauchte Zeit	Bewertung	F 1	F 2	F 3	F 4	F 5	F 6	F 7	F 8	F 9	F 10
1	Dieter	huth@illias.de	Beendet	8. Oktober 2012 09:05	8. Oktober 2012 09:44	39 Minuten 44 Sekunden	10	2	2	2	2	2	2	2	2	2	2

Alle auswählen (Alle abwählen) Markierte Versuche neu bewerten Ausgewählte Versuche löschen

Seite: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 (Weiter)

Gesamtzahl der Teilnehmerinnen, die einzelne Bewertungspunkte erreicht haben

Abbildung 19: Moodle: Test-Bericht

Auch LPLUS bietet umfangliche Prüfungsprotokolle:

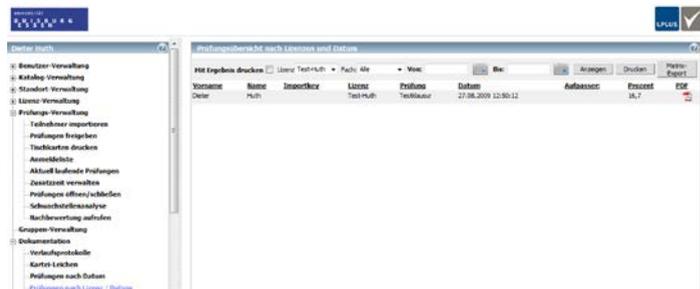


Abbildung 20: LPLUS: Prüfungsübersicht

Ihr Ergebnis: 16.50 von 64.00 Punkte (25.78 %)

Thema	Anzahl der Aufgaben	Erreichte Punkte	Maximale Punkte	Prozent
1. Zwei Paradigmen politischer Theorie	4	4.00	8.50	47.06
2. Grundbegriffe: Gemeinwohl/Gerechtigkeit	3	5.50	11.00	50.00
3. Berlin	1	0.00	2.00	0.00
4. Friedman	2	1.50	5.00	30.00
5. Nozick	2	0.00	6.00	0.00
6. Rawls	2	0.00	6.00	0.00
7. Sandel	2	0.00	5.00	0.00
8. Taylor	2	0.00	6.00	0.00
10. Etzioni	2	1.50	5.00	30.00
11. Barber	2	1.50	5.00	30.00
12. Rawls2	2	2.50	4.50	55.56
Summe	24	16.50	64.00	25.78

Abbildung 21: LPLUS: Teilnehmerübersicht

Anlage 3: Ergebnisübersicht pro Frage

Nummer	interne Nummer	Struktur Kapitel	Punkte	Max. Punkte
1	11	2. Grundbegriffe: Gemeinwohl/Gerechtigkeit	0.00	3.00
2	1	1. Zwei Paradigmen politischer Theorie	0.00	2.00
3	7	5. Nozick	0.00	3.00
4	15	10. Etzioni	1.50	2.00

Abbildung 22: LPLUS: Ergebnisübersicht pro Frage

Statistische Auswertungen

Über einen Filter können Sie sich bestimmte Evaluationsdaten für einzelne oder mehrere Teilnehmer ansehen und als XLS-Datei exportieren. Über die Funktion "De-
taillierte Statistik anzeigen", lassen sich weitere Details zu den Ergebnissen eines Teilnehmers finden.

The screenshot shows the ILIAS evaluation interface. At the top, there are numbered tabs: 2., 1., 3., 5., 6., 7., 4. Below these are navigation tabs: Fragen, Info, Einstellungen, Teilnehmer, Manuelle Bewertung, Statistik, Verlauf, Metadaten, Export, Rechte. The main content area has tabs for 'Auswertung für alle NutzerInnen', 'Aggregierte Testergebnisse', and 'Ergebnisse zu Einzelfragen'. There are options to export data as 'Microsoft Excel' and an 'Export' button. A filter section allows filtering by Name, Gruppe, Kurs, and 'Nur bestandene Tests'. Below the filter is a table with columns: Name, Benutzername, Erreichte Punkte, Note, Beantwortete Fragen, Bearbeitungszeit, and 'Detaillierte Statistik'. The table shows two rows of data. The 'Detaillierte Statistik' column contains a button 'Detaillierte Statistik anzeigen' which is highlighted with a red box. A blue arrow points from the text above to this button.

Name	Benutzername	Erreichte Punkte	Note	Beantwortete Fragen	Bearbeitungszeit	Detaillierte Statistik
[]	[]	4 von 11	nicht bestanden	7 von 7 (100,00 %)	00:01:59	Detaillierte Statistik anzeigen
[]	[]	0 von 12	nicht bestanden	5 von 7 (71,43 %)	00:00:20	Detaillierte Statistik anzeigen

Abbildung 23: ILIAS: Evaluationsdaten (Entnommen: KIM IT-Dienste der Universität Hohenheim, wie erstelle ich Tests in ILIAS (Version 4.2))

Moodle Test-Information

The screenshot shows the Moodle Test-Information page. At the top, there are settings for the page: 'Einstellungen für diese Seite' and 'Statistik berechnen aus' with a dropdown menu set to 'Alle Versuche'. There is a 'Bericht anzeigen' button. Below this is the 'Test-Information' section. It includes a 'Bericht herunterladen als' button and a dropdown menu set to 'Textdatei mit kommagetrennten Werten (CSV)'. The main content is a table with two columns: 'Test-Name' and 'Abschusstest Kapitel 1'. The table shows various statistical metrics for the test 'UNITS-01 Grundlagen der Mathematik (WS 2012/13)'. A blue arrow points from the text above to the 'Bericht anzeigen' button.

Test-Name	Abschusstest Kapitel 1
Kursname	UNITS-01 Grundlagen der Mathematik (WS 2012/13)
Anzahl der vollständig bewerteten ersten Versuchen	3
Gesamtzahl an vollständig bewerteten Versuchen	5
Durchschnitt bei erstem Versuch	73,61%
Durchschnittsergebnis aller Versuche	61,94%
Median (für Alle Versuche)	50,00%
Standardabweichung (für Alle Versuche)	29,68%
Schiefe der Punkteverteilung (für Alle Versuche)	0,4428
Bewertungsverteilungsgraph (für Alle Versuche)	-3,0539
Koeffizient interner Konsistenz (für Alle Versuche)	67,72%
Fehlerquotient (für Alle Versuche)	58,81%
Standardfehler (für Alle Versuche)	16,86%

Abbildung 24: Moodle: Test-Information

Test-Strukturanalyse

Tabellendaten laden als | Testdaten mit normierungsbereinigten Werten (G:0%)

Fr.	Titel der Frage	Versuche	Möglichkeitstermin	Standardabweichung	Zufällig angrenzende Punktezahl	Bearbeitungszeit	Gewichtung	Effektive Gewichtung	Diskriminationsindex	Unterschiedliche Effizienz
1	Quiz (Zahlenreihen und Sternkopierer)	5	80,00%	44,72%			16,67%	19,25%	-2,68%	-4,35%
2	Quiz (Zahlenreihen und Sternkopierer)	6	70,00%	29,61%			16,67%	9,40%	12,06%	12,68%
	4+3 - Polarkoordinaten	1	100,00%		16,67%		16,67%			
	Rechenoperationen bei natürlichen Zahlen	2	100,00%	0,00%			16,67%			
	Rechenoperationen bei ganzen Zahlen	1	100,00%				16,67%			
	(10) - (50)	1	100,00%		33,33%		16,67%			
	Zu welchem Zahlenbereich gehört 8896,5	2	80,00%	23,57%			16,67%		-100,00%	-100,00%
	ist „5 eine natürliche Zahl“	1	0,00%		50,00%		16,67%			
	Ausdr. / Substitution formaler	1	100,00%		50,00%		16,67%			
	Wahrheit von 4x	1	80,00%		0,00%		16,67%			

Abbildung 25: Moodle: Test-Strukturanalyse

Statistische Auswertungen in LPLUS

Teilnehmeranzahl

Teilnehmer gesamt:40

Notenspiegel

Note	Text	Häufigkeit der Note	Prozent
1	sehr gut	5	12.50 %
1,3	sehr gut	6	15.00 %
1,7	gut	8	20.00 %
2	gut	5	12.50 %
2,3	gut	3	7.50 %
2,7	befriedigend	2	5.00 %
3	befriedigend	2	5.00 %
3,3	befriedigend	4	10.00 %
3,7	ausreichend	1	2.50 %
4	ausreichend	2	5.00 %
5	nicht bestanden	2	5.00 %

Notendurchschnitt:

Gewichteter Notendurchschnitt:2,24

Modalwert (häufigste Note):1,7 (gut)

Variationskoeffizient

Der Variationskoeffizient beträgt:0,11

Abbildung 26: LPLUS: Notenspiegel, Notendurchschnitt, Variationskoeffizient

Vergleich der gängigen Prüfungssoftwarelösungen

Übersicht: Schwierigkeitsgrad nach Aufgabe

Interne Nummer	Frageart	Häufigkeit der Note	Prozent
4	Multiple Choice	40	73,93
5	Multiple Choice	40	90,36
6	Multiple Choice	40	84,29
7	Multiple Choice	40	80,83
8	Multiple Choice	40	75,00
9	Multiple Choice	40	92,50
10	Multiple Choice	40	78,50
11	Multiple Choice	40	73,50
12	Multiple Choice	40	
13	Multiple Choice	40	
16	Ungebunden	40	
18	Ungebunden	40	
20	Ungebunden	40	
21	Ungebunden	40	

Übersicht: Trennschärfe

Interne Nummer	Frageart	Häufigkeit der Note	Trennschärfe
4	Multiple Choice	40	0,79
5	Multiple Choice	40	0,78
6	Multiple Choice	40	0,76
7	Multiple Choice	40	0,61
8	Multiple Choice	40	0,69
9	Multiple Choice	40	0,66
10	Multiple Choice	40	0,69
11	Multiple Choice	40	0,53
12	Multiple Choice	40	0,77
13	Multiple Choice	40	0,81
16	Ungebunden	40	0,79
18	Ungebunden	40	0,65
20	Ungebunden	40	0,64
21	Ungebunden	40	0,75

Abbildung 27: LPLUS: Schwierigkeitsgrad, Trennschärfe

Ergebnis: **31,2 % (20 von 64 Punkten erreicht)**

Themenübersicht

Thema	Anzahl Aufgaben	Erreichte Punkte	Maximale Punkte	Prozent
1. Zwei Paradigmen politischer Theorie	4	4,50	8,50	52,94
2. Grundbegriffe: Gemeinwohl/Gerechtigkeit	3	4,50	11,00	40,91
3. Berlin	1	0,00	2,00	0,00
4. Friedman	2	2,50	5,00	50,00
5. Nozick	2	0,00	6,00	0,00
6. Rawls	2	0,00	6,00	0,00
7. Sandel	2	0,00	5,00	0,00
8. Taylor	2	0,00	6,00	0,00
10. Etzioni	2	2,00	5,00	40,00
11. Barber	2	2,00	5,00	40,00
12. Rawls2	2	4,50	4,50	100,00
Summe	24	20,00	64,00	31,25

Abbildung 28: LPLUS: Schwachstellenanalyse

Fazit

Mein Bewertungsansatz war:

- Bei den Bewertungseinstellungen liegen die größten Unterschiede und daraus folgend die eigentliche individuelle Entscheidung für oder gegen ein Prüfungs- / Test-System.

Momentanes Vergleichsergebnis:

- Die Fragetypen werden mehr und die Entwicklung schreitet bei den Open-Source-Systemen dank der aktiven Community schneller voran.
- Die Bewertungseinstellungen der einzelnen Fragetypen und die Prüfungsmanagementumgebung sind meiner Meinung nach bei LPLUS mächtiger.

An der Uni Wuppertal planen wir, sowohl mit E-Prüfungen in der bei uns etablierten LMS-Umgebung Moodle als auch mit LPLUS zu starten und die Nutzungsentcheidung den Prüfungsverantwortlichen nach entsprechenden Schulungen zu überlassen.

Die dabei gemachten Erfahrungen und die weitere Entwicklung der Prüfungssoftware-Lösungen werden das zukünftige Vorgehen beeinflussen und gegebenenfalls korrigieren.

Referenzen

ILIAS: Benutzerdokumentation. Abgerufen unter http://www.ilias.de/docu/goto_docu_lm_3052.html [07.04.2015]

Moodle: Dokumentation. Abgerufen unter <https://docs.moodle.org/26/de/Hauptseite> [07.04.2015]

LPLUS Teststudio. Abgerufen unter <http://www.lplus.de/Manual/de/> [07.04.2015]

Kommunikations-, Informations- und Medienzentrum (KIM) der Universität Hohenheim. Abgerufen unter <https://kim.uni-hohenheim.de/onlineklausuren> [07.04.2015]

Vita

Dieter Huth ist seit dem 01.07.2013 Leiter des Zentrums für Informations- und Medienverarbeitung an der Bergischen Universität Wuppertal. Vor seinem Wechsel nach Wuppertal war er seit 1982 an der Universität Essen und der fusionierten Nachfolgeeinrichtung Universität Duisburg-Essen beschäftigt. Dort verantwortete er ab 2007 den Aufbau eines Kompetenzzentrums für PC-gestützte Prüfungen, in dessen Verlauf Ende 2009 eine Papierklausurenfläche zu einer PC-Hall mit 198 Computerarbeitsplätzen umgebaut wurde. Zu seinen damaligen Aufgaben gehörte auch die technische Mitbetreuung der LPLUS-Prüfungsumgebung.

Automatisierte Bewertung von Programmieraufgaben: Szenarien, Werkzeuge, Perspektiven

Zusammenfassung

Ziel der Programmierausbildung ist es, die Studierenden in die Lage zu versetzen, eigenständig Programmcode zu schreiben, der eine gegebene Aufgabestellung erfüllt. Dazu ist es unerlässlich, dass die Studierenden einerseits regelmäßig üben und dass andererseits ihre Fähigkeiten in der Prüfung angemessen berücksichtigt werden. Um beiden Zielen gerecht zu werden, führt die Universität Duisburg-Essen in der Programmierausbildung der Erstsemester am Campus Essen seit 2007 nicht nur einen klassischen Vorlesungs- und Übungsbetrieb durch, sondern bietet im Semesterverlauf regelmäßige Testate an, in denen eine Mindestpunktzahl erworben werden muss, um an der abschließenden Klausur teilnehmen zu dürfen. Die Studierenden werden so einerseits zu regelmäßiger Übung angehalten und es ist andererseits sichergestellt, dass nur Studierende mit hinreichenden Programmierfähigkeiten die Vorlesung erfolgreich abschließen.

Aufgrund von Teilnehmerzahlen von über 500 Studierenden pro Semester ist es nicht möglich, Übungen und Testate manuell zu korrigieren und zeitnah mit Feedback zu versehen. Daher kommt zur automatischen Bewertung der Programmieraufgaben die Software JACK zum Einsatz, deren Entwicklung ebenfalls von der Universität Duisburg-Essen betrieben wird. Die Kernfunktionalität der Software besteht darin, Lösungen zu Übungs- und Prüfungsaufgaben automatisch nach verschiedenen Kriterien und individuellen Vorgaben je nach Aufgabenstellung zu analysieren und dazu passende Rückmeldungen in Form von Punktzahlen und hilfreichen Feedbackmeldungen zu erzeugen. Die didaktischen und technischen Möglichkeiten des Systems – auch im Vergleich zu anderen derartigen Systemen – sowie der nötige Aufwand zur Vorbereitung geeigneter Aufgaben werden in diesem Beitrag vorgestellt.

Einleitung

Der Einsatz automatisierter Werkzeuge zur Bewertung der Lösungen von Übungsaufgaben in der Informatik und insbesondere in der Programmierausbildung ist nicht nur aufgrund des fachlichen Zusammenhangs naheliegend, sondern auch

einer der ältesten Aspekte von E-Learning überhaupt. Erste derartige Werkzeuge wurden bereits in den 1960er Jahren publiziert (Hollingsworth, 1960; Forsythe & Wirth, 1965). Seitdem wurden zahlreiche neue Werkzeuge entwickelt, um neue Funktionen umsetzen zu können und mit der Entwicklung von Programmiersprachen und der allgemeinen Leistungsfähigkeit von Rechnern mitzuhalten. Unabhängig davon, dass heutige Programmiersprachen komplexer sind als vor 50 Jahren und heutige Rechner deutlich leistungsfähiger, sind die Grundfragen der Forschung und Entwicklung in diesem Bereich noch immer dieselben: Wie kann ein Computer feststellen, ob die Lösung zu einer Übungsaufgabe korrekt ist und wie kann er zu einer teilweise oder vollständig inkorrekten Lösung sinnvolles Feedback geben, das den Studierenden hilft, die Lösung zu verbessern und das eigentliche Lernziel zu erreichen?

Zufriedenstellende Antworten auf diese Fragen sind umso dringender, je mehr Studierende sich für ein Informatikstudium entscheiden und die dazu nötige Programmierausbildung in den ersten Semestern durchlaufen. An der Universität Duisburg-Essen sind am Campus Essen beispielsweise regelmäßig mehr als 500 Teilnehmerinnen und Teilnehmer in der Erstsemestervorlesung *Programmierung* zu verzeichnen. Die Erfahrung hat gezeigt, dass diese durch regelmäßige Aufgaben und Übungen dazu angehalten werden sollten, ihre Programmierfähigkeiten zu trainieren. Aufgrund der großen Zahl von Studierenden ist es dabei unerlässlich, dass eine automatisierte Bewertung von Programmieraufgaben vorgenommen wird, um zeitnah und detailliert Feedback zu den einzelnen Lösungen geben zu können. Dazu kommt das an der Universität Duisburg-Essen entstandene Werkzeug JACK zum Einsatz, das seit 2007 entwickelt wird und über typische Funktionen zur Bewertung von Programmieraufgaben verfügt.

Im Folgenden werden zunächst verschiedene Einsatzszenarien vorgestellt, in denen JACK an der Universität Duisburg-Essen zum Einsatz kommt. Anschließend werden die Funktionen von JACK sowie vergleichbarer Werkzeuge erläutert und der Aufwand für die Lehrenden für den Einsatz diskutiert. Abschließend werden Perspektiven und Grenzen der weiteren Entwicklung in diesem Bereich aufgezeigt.

Szenarien

Grundsätzlich unterscheidet sich der Einsatzbereich von Automatisierungswerkzeugen für die Bewertung von Lösungen nicht von der manuellen Korrektur von Lösungen, so dass im Wesentlichen dieselben Szenarien des Übungs- und Prüfungsbetriebs abgedeckt werden können, beginnend beim freien Übungsbetrieb zum Selbstlernen über verpflichtende Übungsaufgaben bis hin zur summativen Prüfung.

Freier Übungsbetrieb (Self-Assessment)

Im freien Übungsbetrieb dienen JACK oder vergleichbare Systeme den Studierenden als Trainingspartner, der die Rolle eines Tutors übernimmt. Das System stellt dazu Übungsaufgaben zum wahlfreien Zugriff zur Verfügung, die von den Studierenden in beliebiger Reihenfolge und beliebiger Häufigkeit bearbeitet werden können. Jede Einreichung wird durch das System geprüft und die Beurteilung den Studierenden unmittelbar zur Verfügung gestellt. Manuelles Eingreifen durch Lehrende ist in diesem Szenario primär auf Situationen beschränkt, in denen das System mit außergewöhnlichen Lösungen konfrontiert ist. In diesem Sinne dient das System als Filter, das den Lehrenden hilft, sich auf besondere Fälle zu konzentrieren und entlastet sie von der Behandlung vieler gleichartiger Lösungen.

Der Fokus der automatischen Bewertung liegt in diesem Szenario auf der Generierung sinnvollen, primär deskriptiven Feedbacks, das das Verständnis für die Aufgabe und die Schwächen der eingereichten Lösung fördert. Evaluierendes Feedback im Sinne von Punktzahlen kann als motivierender Faktor hinzutreten, insbesondere um einen Fortschritt im Verlauf von mehreren, schrittweise verbesserten Einreichungen zu einer Aufgabe zu verdeutlichen.

Konzeptionell kann in diesem Szenario auch ein adaptives Systemverhalten realisiert werden, indem ein System aus den bisherigen Einreichungen eines Studierenden lernt und weitere passende Aufgaben zur Bearbeitung vorschlägt, die gezielt entdeckte Schwächen ansprechen.

Strukturierte Übung (Formatives Assessment)

Der strukturierte Übungsbetrieb unterscheidet sich vom vorherigen Szenario vor allem durch organisatorische Rahmenbedingungen, durch die bestimmte Aufgaben für die Bearbeitung in einem bestimmten Zeitraum vorgeschrieben sind und als Leistungskontrolle dienen. Am Campus Essen der Universität Duisburg-Essen wird dieser Modus im Laufe eines Semesters für sechs sogenannte Miniprojekte eingesetzt, für die jeweils eine feste Frist für die Abgabe existiert. Die Bearbeitung dieser Aufgaben ist – unabhängig vom erreichten Ergebnis – Voraussetzung für die Teilnahme an regelmäßig stattfindenden Testaten. Wie im freien Übungsbetrieb ist manuelles Eingreifen durch Lehrende in diesem Szenario nur notwendig, wenn es zu besonderen Situationen kommt.

Auch wenn das konkrete Ergebnis für die Zulassung zum Testat keine Rolle spielt, liegt in diesem Szenario ein stärkerer Fokus auf dem evaluierenden Feedback als im freien Übungsbetrieb. Die erreichte Punktzahl wird insbesondere von den Studierenden als Leistungsindikator aufgefasst, der Auskunft darüber gibt, mit welchem Ergebnis sie im Testat rechnen können. Nicht schlüssige Bewertungen werden dementsprechend auch als unfair wahrgenommen. Gleichzeitig spielt das deskripti-

ve Feedback eine wichtige Rolle, da es als weiterer Leistungsindikator dient und insbesondere auch zum Vergleich zwischen den Studierenden genutzt wird.

Konzeptionell kann in diesem Szenario ebenfalls adaptives Systemverhalten realisiert werden, indem die ermittelten Punktzahlen zur Anpassung der Schwierigkeit von nachfolgenden Aufgaben genutzt werden und es den Studierenden so erspart wird, Aufgaben lösen zu müssen, die sie deutlich unterfordern.

Prüfungsbetrieb (Summatives Assessment)

Summatives Assessment wird in der Erstsemestervorlesung *Programmierung* in Form von Testaten durchgeführt, die sechsmal im Semester angeboten werden. In diesen Testaten müssen die Studierenden einzelne Aufgaben innerhalb von 45 Minuten bearbeiten und erhalten in dieser Zeit kein Feedback durch JACK. In diesem Szenario übernimmt JACK daher ausschließlich die Rolle des Korrektors, der die eingereichten Lösungen mit einer Punktzahl versieht. Aufgrund des Prüfungscharakters dieses Szenarios ist hier bereits aus rechtlichen Gründen eine manuelle Überwachung notwendig, mit der alle Ergebnisse der automatischen Prüfung zunächst freigegeben werden müssen, bevor die Resultate weiter verwendet werden, um aus der erreichten Punktzahl in den Testaten die Zulassung zur Abschlussklausur zu bestimmen. Grundsätzlich kann auch in dieser JACK im selben Modus eingesetzt werden, was bisher jedoch aus organisatorischen Gründen an der Universität Duisburg-Essen nicht passiert ist, da die Klausur nicht nur aus Programmieraufgaben besteht und die Lehrenden eine Hybrid-Klausur vermeiden wollten.

Der Fokus der Arbeit von JACK liegt in diesem Szenario damit klar auf einem gezielten, möglichst detaillierten evaluierenden Feedback in Form von Punktzahlen. Deskriptives Feedback tritt in den Hintergrund und dient lediglich zur Erläuterung der erreichten Punktzahl, wenn die Studierenden im Sinne einer Prüfungseinsicht ihre Lösungen und die Bewertung im Nachhinein einsehen.

Werkzeuge und Funktionen

Die Generierung von Feedback erfolgt in JACK auf verschiedenen Wegen, so dass auch verschiedene Formen von Feedback möglich sind. Vergleichbare Systeme bieten zum Teil ähnliche und zum Teil komplementäre Funktionen an, wobei es eine gewisse Schnittmenge gibt, die als Standardfunktionalität angesehen werden kann. Insgesamt gibt es eine sehr große Anzahl von Systemen, die die automatische Bewertung von Programmieraufgaben in verschiedenen Programmiersprachen anbieten. Für einen Überblick wird an dieser Stelle auf die Reviews von Ihantola et al. (2010) sowie Ala-Mutka (2005) verwiesen.

Standardfunktionen

Grundsätzlich besteht die Möglichkeit, Programmcode nach statischen und dynamischen Gesichtspunkten zu prüfen. Eine statische Prüfung betrachtet dabei den Programmcode als solchen und untersucht ihn auf erwünschte und unerwünschte Eigenschaften (Striewe & Goedicke, 2014), während eine dynamische Prüfung den Programmcode zur Ausführung bringt und die dabei erzeugten Ausgaben untersucht.

Wichtigstes Werkzeug der statischen Prüfung ist der Compiler, der auf syntaktische Fehler hinweist. Zur weiteren Analyse betrachtet JACK den Syntaxgraphen des Programms und kann dort beispielsweise prüfen, ob eine Aufgabe mit einer iterativen Strategie durch den Einsatz von Schleifenkonstrukten gelöst wurde oder mit einer rekursiven Strategie. Je nach Aufgabenstellung kann dies als Fehler oder aber auch als besonders positiver Aspekt einer Lösung gewertet und im Feedback entsprechend berücksichtigt werden. Ferner ist auf diese Weise in beschränktem Rahmen eine stilistische Prüfung möglich, indem beispielsweise die Einhaltung von Konventionen zur Groß- und Kleinschreibung von Bezeichnern geprüft wird. Die Berücksichtigung optischer Aspekte wie beispielsweise die Verwendung von Einrückungen und Leerzeilen an geeigneten Stellen ist dagegen auf diese Weise nicht möglich und wird daher von JACK nicht angeboten. Andere Systeme wie beispielsweise der Praktomat setzen für derartige Überprüfungen externe Werkzeuge wie CheckStyle ein, die auch im industriellen Kontext Verwendung finden (Krinke, Störzer, & Zeller, 2002). Weitere externe Werkzeuge für die statische Analyse sind FindBugs und PMD, die beispielsweise von ASB (Morth, Oechsle, Schloß, & Schwinn, 2007) oder DUESIE (Hoffmann, Quast & Wismüller, 2008) eingesetzt werden. Nicht immer sind diese Werkzeuge für die Zwecke eines E-Learning-Systems geeignet oder sie lassen sich nur mit einem gewissen Aufwand passend einrichten.

Die dynamische Prüfung einer Lösung geschieht in der Regel durch das Ausführen von Testfällen, die Eingaben an den Programmcode senden und die Ausgabe mit einer Vorgabe vergleichen. Abweichungen werden als fehlgeschlagene Testfälle gewertet und können entsprechend als Feedback ausgegeben werden. In JACK erfolgt die Durchführung von Testfällen für die Programmiersprache Java durch ein eigenes Testsystem, während bei vielen anderen Systemen der Einsatz des externen Werkzeugs JUnit üblich ist. Die vergleichbare Variante CPPUnit für die Programmiersprache C++ kommt auch in JACK zum Einsatz (Hesse, Wagner, & Paech, 2013). Ein wichtiges Unterscheidungsmerkmal zwischen verschiedenen Ansätzen ist dabei, ob lediglich Ergebnisse verglichen werden, oder ob die gesamte Testausführung beobachtet und aufgezeichnet wird. Letzteres geschieht in JACK bei Java-Aufgaben in Form von sogenannten Traces, in denen der komplette Programmablauf schrittweise nachvollzogen werden kann. Die Traces tragen nicht zur Bestim-

mung einer Punktzahl bei, helfen den Studierenden aber, ihre eigene Lösung besser zu verstehen und gegebenenfalls auch Fehler zu identifizieren (Striewe & Goedicke, 2011).

Besondere Funktionen

Durch die oben genannte Aufzeichnung aller Ausführungsschritte einer Lösung ist es in JACK zudem möglich, Zusatzfunktionen anzubieten, die insbesondere die Bereitstellung von deskriptivem Feedback betreffen. So wird den Studierenden angezeigt, welche Zeilen ihres Programmcodes von keinem der Testfälle betroffen waren und somit möglicherweise für die Erfüllung der Aufgabe irrelevant sind. Ferner kann ein automatischer Abgleich der Traces mit dem Trace einer Musterlösung vorgenommen werden, um bei fehlerhaften Lösungen genauer einzugrenzen, welcher Teil des Programmcodes den Fehler verursacht (Striewe & Goedicke, 2013). Schließlich kann auf Basis der Aufzeichnungen auch eine Visualisierung von Datenstrukturen erstellt werden, mit der die Studierenden einen Programmzustand in einer Notation betrachten können, die ihnen beispielsweise aus der Vorlesung bereits geläufig ist (Striewe & Goedicke, 2010).

Weitere besondere Funktionen findet man insbesondere in Werkzeugen, die bestimmte Szenarien unterstützen. Beispielsweise fokussiert das System GATE auf einen prüfungsorientierten Einreichungsprozess und bietet dazu unter anderem eine automatische Plagiatskontrolle an (Strickroth, Olivier, & Pinkwart, 2011). Diese ermöglicht es den Lehrenden, verdächtige Einreichungen schnell zu identifizieren und die betroffenen Studierenden entsprechend zu kontaktieren.

Datenaustausch und Integration

Angesichts der Vielzahl an existierenden Werkzeugen und ihrer Funktionen stellt sich zwangsläufig die Frage, ob Aufgaben zwischen verschiedenen Werkzeugen ausgetauscht werden können. Dazu wird derzeit nur sehr wenig Unterstützung angeboten. Es existiert ein Austauschformat, das mit den Systemen JACK, GATE und Praktomat genutzt werden kann und das insbesondere berücksichtigt, dass bestimmte systemspezifische Überprüfungen auf anderen Systemen nicht zur Verfügung stehen (Strickroth et al., 2015).

Verbreiteter ist die Möglichkeit der Integration von automatischen Prüfungssystemen in Learning Management Systeme (LMS) oder vergleichbare Umgebungen. Über den IMS-LTI-Standard ermöglicht JACK die Integration in alle LMS, die diesen Standard ebenfalls implementieren. Da der Standard nur eine sehr schmale Kommunikation zwischen LMS und integriertem Werkzeug ermöglicht, arbeiten Becker et al. (2013) an weiteren Integrationsmöglichkeiten. Andere Systeme wie beispielsweise eduComponents (Amelung, Forbrig & Rösner, 2008) sind gleich gar

nicht als eigenständige Werkzeuge entworfen worden, sondern wurden von Anfang an für die Integration in LMS oder Content Management Systeme konstruiert.

Aufwand

Regelmäßig wird mit dem Einsatz automatischer Bewertungssysteme das Versprechen verbunden, dass sich der Aufwand für die Lehrenden erheblich reduziert, weil das zeitintensive Korrigieren von Lösungen entfällt. Gleichzeitig wird ein erheblich höherer Vorbereitungsaufwand für die Erstellung von Aufgaben angeführt. Beide Beobachtungen lassen sich durch den langjährigen Einsatz von JACK an der Universität Duisburg-Essen sowie an weiteren Universitäten und Hochschulen bestätigen.

Um eine Aufgabe vollständig vorzubereiten, müssen zunächst wie bei herkömmlichen Fragestellungen ein Aufgabentext sowie eine Musterlösung entworfen werden. Letztere ist für die Grundfunktionalität nicht zwingend erforderlich, ermöglicht aber die leichte Überprüfung der weiteren Vorbereitungen. Idealerweise werden sogar mehrere Musterlösungen angefertigt oder auch prototypische fehlerhafte Lösungen, um das Systemverhalten für diese Fälle überprüfen zu können. Anschließend müssen für alle gewünschten Prüfverfahren (statische und dynamische Überprüfung) die nötigen Testfälle, Prüfregeln, usw. inklusive zugehöriger Feedbacknachrichten erstellt werden. Diese Arbeit macht einen erheblichen Teil der Vorbereitungsleistung aus und erfordert wesentlich mehr Präzision als vergleichbare Vorüberlegungen bei manuell korrigierten Aufgaben. Insbesondere bestimmt hier auch der Lernaufwand für die Nutzung der jeweiligen Notationen für Testfälle und Prüfregeln die Lernkurve beim erstmaligen Einsatz automatischer Prüfungssysteme. Der anschließende Aufwand für die Erstellung von Punkteschemata bzw. Gewichtungen der einzelnen Prüfverfahren entspricht dagegen wiederum dem, was auch bei manuell korrigierten Aufgaben vorab zu leisten ist. Insgesamt kann aus der Erfahrung mit JACK von einem Aufwand von bis zu 20 Stunden bei der erstmaligen Arbeit mit JACK sowie von im Schnitt fünf bis acht Stunden für die Einrichtung einer Aufgabe gesprochen werden, wobei in diesen Zeiten der Bedarf für den eigentlichen Entwurf der Aufgabe nicht berücksichtigt ist.

Diesem Aufwand steht die erhoffte Zeitersparnis bei der Nachbereitung der Aufgaben gegenüber, denn erfahrungsgemäß muss auch für die stichprobenartige Kontrolle von 1.000 Einreichungen im Übungsbetrieb nicht mehr als eine Stunde aufgewendet werden. Auch die manuelle Abschlusskontrolle bei Testaten benötigt kaum mehr Zeit, so dass mit einem Aufwand von fünf bis zehn Stunden pro Testat gerechnet werden kann.

Langfristig ist zudem zu berücksichtigen, dass Aufgaben problemlos wiederverwendet werden können und in solchen Fällen lediglich Aufwand für die Qualitätssi-

cherung betrieben werden muss, d. h. insbesondere für die Verbesserung der Feedbackmeldungen zur Berücksichtigung weiterer typischer Fehlerfälle. Der Austausch von Aufgaben zwischen Nutzern desselben Systems oder über Systemgrenzen hinweg ermöglicht es zudem, einen großen Aufgabenpool zusammenzustellen und gemeinsam weiterzuentwickeln, was den Aufwand für den Einzelnen noch einmal senkt.

Perspektiven und Grenzen

Perspektiven für den weiteren zukünftigen Einsatz automatischer Bewertungssysteme für Programmieraufgaben ergeben sich vor allem daraus, dass sich viele Analyseaufgaben gut automatisieren lassen. Ist eine bestimmte Analyse erst einmal technisch umgesetzt und für eine konkrete Aufgabe konfiguriert, liefern automatische Systeme wesentlich schneller und präziser Feedback an die Studierenden, als dies mit einer manuellen Korrektur möglich wäre. Zum einen ist dabei insbesondere auch zu berücksichtigen, dass in der industriellen Softwareentwicklung zum Teil schon Verfahren existieren, die in E-Learning-Systemen noch nicht zum Einsatz kommen. Zum anderen lassen sich viele Tätigkeiten, die ein Lehrender bei der manuellen Korrektur einer Lösung durchführen würde, mindestens mit heuristischen Verfahren automatisieren. Beispielsweise könnte ein Lehrender zur Sicherstellung einer gerechten Bewertung berücksichtigen, ob ein konkreter Fehler weitreichende Konsequenzen hat oder seine Eliminierung eine inkorrekte Lösung korrekt macht. Ein automatisches System könnte solche Überprüfungen automatisieren, indem es systematisch Teile einer studentischen Lösung durch passende Ausschnitte einer Musterlösung ersetzt.

Grenzen hat der Einsatz automatischer Bewertungssysteme immer dort, wo die kreative Leistung eines Studierenden wichtiger ist als die Umsetzung formaler Regeln. So kann eine automatisierte Bewertung zwar sowohl auf syntaktische als auch inhaltliche Fehler hinweisen und dabei durch den Vergleich mit Musterlösungen möglicherweise sehr präzise benennen, an welcher Stelle ein Fehler liegt, aber dies bedeutet nicht zwangsläufig, dass der Fehler nur durch eine Korrektur an dieser Stelle behoben werden kann. Stattdessen ermöglichen Programmieraufgaben im Allgemeinen beliebig viele korrekte Lösungen und damit auch beliebig viele Möglichkeiten, einen gegebenen Fehler zu beheben. Dies zeigt sich auch regelmäßig in Umfragen unter den Nutzern, in denen die Aussage „JACK kann Fehler mindestens genauso gut finden wie ein Mensch“ mehrheitlich Zustimmung erhält, während die Aussage „JACK kann Fehler mindestens genauso gut erklären wie ein Mensch“ mehrheitlich Ablehnung erfährt.

Fazit

Automatische Bewertungssysteme für Programmieraufgaben bieten eine breite Spanne von didaktischen Szenarien und technischen Möglichkeiten, durch die sie einen Mehrwert für die Lehre darstellen können. Sowohl die Generierung von deskriptivem als auch evaluierendem Feedback ist möglich und kann mit akzeptablem Aufwand konfiguriert werden. Dazu stehen verschiedene Systeme zur Verfügung, die unterschiedliche Schwerpunkte setzen und besondere Funktionen anbieten. Es kann erwartet werden, dass weitere Entwicklungen in der Zukunft noch mehr Funktionen zur Feedbackerzeugung hervorbringen und der Einsatz automatischer Bewertungssysteme für Programmieraufgaben daher noch weiter intensiviert werden kann. Dabei muss jedoch stets berücksichtigt werden, dass sich kreative Leistung, wie sie beim Lösen von Programmieraufgaben notwendig ist, nie vollständig automatisiert unterstützen lässt.

Referenzen

- Ala-Mutka, K. M. (2005). A survey of automated assessment approaches for programming assignments. *Computer science education*, 15(2), S. 83-102.
- Amelung, M., Forbrig, P. & Rösner, D. (2008). Towards generic and flexible web services for e-assessment. In J. Amillo, C. Laxer, E. M. Ruiz & A. Young (Hrsg.), *ITiCSE '08: Proceedings of the 13th annual conference on Innovation and technology in computer science education* (S. 219-224). New York: ACM.
- Becker, S., Stöcker, A., Bräckelmann, D., Garmann, R., Grzanna, S., Heine, F., Kleiner, C., Werner, P. & Bott, O. J. (2013). Prototypische Integration automatisierter Programmbewertung in das LMS Moodle. In U. Priss & M. Striewe (Hrsg.), *Proceedings of the first workshop: „Automatische Bewertung von Programmieraufgaben“ (ABP 2013)*. Hannover: CEUR-WS.
- Forsythe, G. E. & Wirth, N. (1965). Automatic grading programs. *Communications of the ACM*, 8(5), 275-278.
- Hesse, T., Wagner, A. & Paech, B. (2013). Automated assessment of C++ programming exercises with unit tests. In U. Priss & M. Striewe (Hrsg.), *Proceedings of the first workshop „Automatische Bewertung von Programmieraufgaben“ (ABP 2013)*. Hannover: CEUR-WS.
- Hoffmann, A., Quast, A. & Wismüller, R. (2008). Online-Übungssystem für die Programmierausbildung zur Einführung in die Informatik. In S. Seehusen, U. Lucke & S. Fischer (Hrsg.), *DeLFI 2008: Die 6. E-Learning Fachtagung Informatik* (S. 173-184). Bonn: Gesellschaft für Informatik e. V.

Hollingsworth, J. (1960). Automatic graders for programming classes. *Communications of the ACM*, 3(10), 528-529.

Ihantola, P., Ahoniemi, T., Karavirta, V. & Seppälä, O. (2010). Review of recent systems for automatic assessment of programming assignments. In C. Schulte & J. Suhonen (Hrsg.), *Proceedings of the 10th Koli Calling International Conference on Computing Education Research* (S. 86-93). New York: ACM.

Krinke, J., Störzer, M. & Zeller, A. (2002). Web-basierte Programmierpraktika mit Praktomat. *Workshop Neue Medien in der Informatik-Lehre*, 48-56.

Morth, T., Oechsle, R., Schloß, H. & Schwinn, M. (2007). Automatische Bewertung studentischer Software, Workshop „Rechnerunterstütztes Selbststudium in der Informatik“. In Bonn Gesellschaft für Informatik e. V., Eibl, C., Magenheim, J. & Wessner, M. (Hrsg.), *DeLFI 2007: Die 5. E-Learning Fachtagung Informatik e. V.* (S. 109-116). Berlin: Logos.

Strickroth, S., Olivier, H. & Pinkwart, N. (2011). Das GATE-System: Qualitätssteigerung durch Selbsttests für Studenten bei der Onlineabgabe von Übungsaufgaben? In S. Friedrich, A. Kienle & H. Rohland (Hrsg.), *DeLFI 2011: Die 9. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e. V.* (S. 115-126). Dresden: TUDpress.

Strickroth, S., Striewe, M., Müller, O., Priss, U., Becker, S., Rod, O., Garmann, R., Bott, J. O. & Pinkwart, N. (2015). ProFormA: An XML-based exchange format for programming tasks. *eLeed*, 11.

Striewe, M. & Goedicke, M. (2010). Visualizing data structures in an e-learning system. *Proceedings of the 2nd International Conference on Computer Supported Education (CSEDU)*, 1, 172-179.

Striewe, M. & Goedicke, M. (2011). Using run time traces in automated programming tutoring. In G. Rößling, T. L. Naps & C. Spannagel (Hrsg.), *Proceedings of the 16th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE 2011)*, S. 303-307, New York: ACM.

Striewe, M. & Goedicke, M. (2013). Trace alignment for automated tutoring. *Proceedings of International Computer Assisted Assessment (CAA) Conference 2013*.

Striewe, M. & Goedicke, M. (2014). A review of static analysis approaches for programming exercises. In M. Kalz & E. Ras (Hrsg.), *Computer assisted assessment — Research into e-assessment: Proceedings of the 17th International Conference on Computer Assisted Assessment* (1. Aufl., Bd. 439, S. 100-113). Heidelberg: Springer.

Vita

Michael Striewe studierte Informatik an der Technischen Universität Dortmund und Klassische Archäologie an der Ruhr-Universität Bochum. Seit seinem Diplomabschluss in Informatik im Jahr 2007 arbeitet er an der Universität Duisburg-Essen in der Arbeitsgruppe „Spezifikation von Software-Systemen“, wo er 2014 promovierte. Sein Forschungsschwerpunkt sind automatisch Prüfungssysteme für komplexe Aufgabentypen, insbesondere im Bereich der Informatikausbildung.

Potenziale neuer Fragetypen für die Naturwissenschaften

Zusammenfassung

An der MLU sind im Laufe des Jahres 2014 zwei neue Fragetypen als Plugin für die Lernplattform ILIAS entstanden. Das erste Plugin bietet einen Zeichenbereich, mit dem Skizzen und Markierungen auf einem optionalen Hintergrundbild eingetragen werden können. Der zweite Fragetyp ermöglicht über einen speziellen Editor die Erstellung chemischer Formeln und Reaktionen, die automatisiert auswertbar sind. Für das Jahr 2015 sind weitere Fragetypen in Planung und Entwicklung.

Einleitung

Seit zwei Jahren können an der MLU und der Universität Leipzig betreute E-Klausuren mit ILIAS durchgeführt werden. Über 1800 Studierende der MLU, vorrangig aus dem Bereich der Naturwissenschaften, und mehr als 2600 Prüflinge in Leipzig, hier insbesondere aus den Erziehungswissenschaften, haben bereits mindestens eine E-Klausur durchlaufen. Die Prüfungsdurchgänge brachten von den Geprüften und den Prüfenden einige Veränderungswünsche hervor. Manche davon waren nur kleine Anpassungen der Plattform, beispielsweise eine Uhrzeitanzeige während der Klausur, mehr Formatierungsmöglichkeiten für das Eingabefeld bei Freitextaufgaben oder eine Anwerthistorie im Klausurprotokoll der Plattform.

Größerer Weiterentwicklungsbedarf erwuchs aus der Abbildung alter Fragen der Papierklausuren auf die verfügbaren Fragetypen des Prüfungssystems. Ein Großteil des Altbestandes setzt sich aus Freitextaufgaben sowie Variationen von Mehrfachauswahlaufgaben zusammen. Teilweise sind Abbildungen oder Ton- und Videodateien einzubinden, die früher mittels CD-Player oder Projektor abgespielt worden wären. Gegebenenfalls mit Einbußen lässt sich letztlich jede dieser Aufgaben auf die vorhandenen Aufgabentypen abbilden. In einen Problembereich gelangt die Fragentransformation, sobald multimediale Inhalte interaktiv zu nutzen sind. Die Bewertung sollte wenn möglich auch automatisiert erfolgen. Doch was ein Mensch mit einem Blick feststellen kann, ist für Computer bisweilen eine komplexe Aufgabe.

ILIAS ist zur Gänze quelloffen, frei nutzbar und durch Plugins erweiterbar. Die Wünsche der Lehrenden und Studierenden konnten mit vertretbarem Zeitaufwand universitätsintern umgesetzt werden. Das LLZ möchte damit E-Klausuren durch neue beziehungsweise veränderte Funktionen besser in der Hochschule verankern und den Prüfern Impulse für die Fragenerstellung bieten. Das erste Plugin – die *Zeichenaufgabe* – sollte zunächst die grundlegenden Möglichkeiten zur Gestaltung von Bildern bieten und wurde mit dem *Moleküleditor* spezialisiert.

Zeichenaufgabe

Es kann kaum die Rede von E-Klausuren sein, wenn für Markierungen und Skizzen noch Papier und Stift ausgelegt werden müssen. Seit Dezember 2013 kann die Zeichenaufgabe (Abbildung 1) dieses Problem aus der Welt schaffen. Sie eignet sich beispielsweise für das:

- Erstellen von Skizzen
- Ergänzen von (Schrift-)Zeichen
- Einzeichnen von Kurvenverläufen
- Markieren von Bereichen

Zeichnen

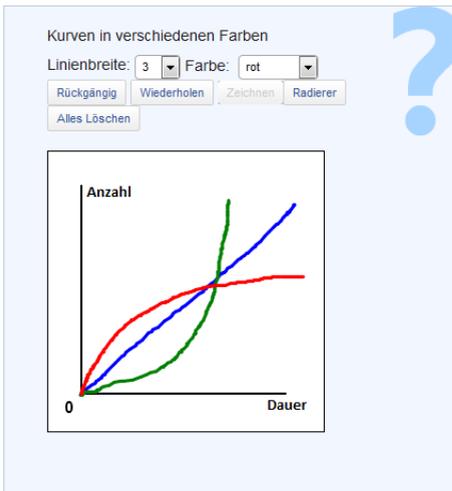


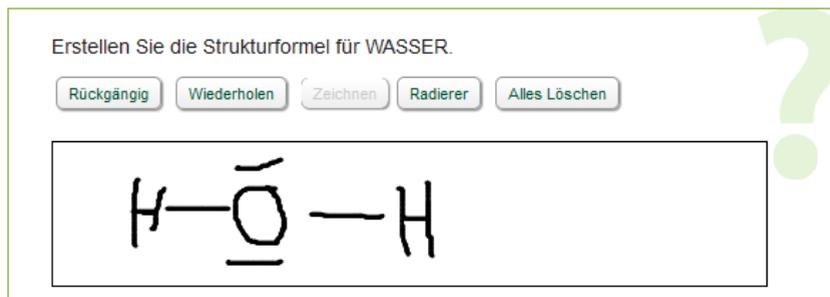
Abbildung 1: Nutzeroberfläche der Zeichenaufgabe

Der Prüfling kann mit verschiedenen Strichstärken und Farben auf einem Zeichenbereich mit optionalem Hintergrundbild arbeiten. Rudimentäre Editorfunktionen (Aktion rückgängig machen, Aktion wiederholen, Radierer, komplette eigene Zeichnung löschen) stehen ebenfalls zur Verfügung. Strichstärken- und Farbauswahl können bei der Fragenerstellung explizit aktiviert oder deaktiviert werden. Auch die Größe des Zeichenbereiches ist einstellbar.

Eine automatisierte Korrektur der Zeichnung erfolgt bislang nicht, da die Aufgabe dem Prüfling viele Freiheitsgrade lässt und der Bildvergleich mit einem Muster für eine klare Bewertung viele Unwägbarkeiten abzudecken hätte. Ein Bewertungsalgorithmus müsste etwa verschiedene Bewertungsgranularitäten besitzen, wie die Punktevergabe in Abhängigkeit einer bestimmten Zeichnungsstruktur, die an einer bestimmten Stelle befindlich sein muss und womöglich in Relation zu anderen Zeichnungen auf der Arbeitsfläche steht. Ein simpler Bildvergleich wird dennoch experimentell in den Aufgabentyp integriert, um zumindest für mit geringem Zeichenaufwand lösbare Aufgaben eine Bewertung auf Basis eines Ähnlichkeitsmaßes zu erstellen und erst bei zu hoher Abweichung die eingesandte Lösung der manuellen Korrektur zuzuführen. Eine andere angedachte Weiterentwicklung des Aufgabentyps ist die Modernisierung der Oberfläche zur besseren Verwendbarkeit von Touch-Eingabegeräten.

Der Aufgabentyp scheint geeignet, den Großteil der bildbasierten Aufgaben von Papierklausuren abbilden zu können. Die Verwendbarkeit hat aber auch Grenzen. Würde die Zeichenaufgabe etwa in einer Prüfung zur Erstellung chemischer Formeln genutzt werden, könnte es ähnlich wie in Abbildung 2 aussehen.

Chemische Formeln



Erstellen Sie die Strukturformel für WASSER.

Rückgängig Wiederholen Zeichnen Radierer Alles Löschen

H— $\overset{\cdot\cdot}{\underset{\cdot\cdot}{\text{O}}}$ —H

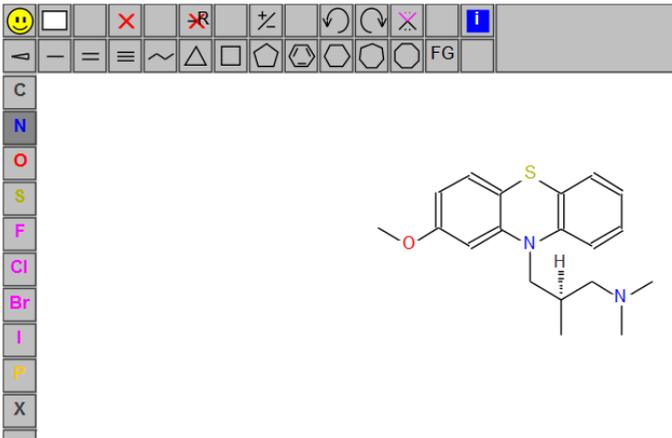
Abbildung 2: Chemische Formeln mit der Zeichenaufgabe

Diese Verwendung erweist sich als wenig praktikabel. Die Buchstaben sind ohne Tablet mit Stifteingabe umständlich zu zeichnen, eine automatische Auswertung gibt es nicht. Diese Herausforderung konnte mit einer Spezialisierung der Zeichenaufgabe gelöst werden.

Moleküleditor

Für die Lernplattform Moodle gibt es seit Juni 2013 ein Fragetyp-Plugin, das die Erstellung von Molekülen über ein Java Applet, den *Java Molecule Editor (JME)*, ermöglicht. Dieses war der Ideengeber für das im Juni 2014 veröffentlichte ILLIAS-Plugin der MLU (Abbildung 3). Da die Applet-Technologie inzwischen überholt ist, sollte jedoch eine JavaScript-basierte Lösung genutzt werden. Es gibt eine kleine Auswahl solcher Editoren, die ohne zusätzliche Programme direkt im Browser ausgeführt werden können. Einer davon ist der *JavaScript Molecule Editor (JSME)*, der Nachfolger des JME. Er steht unter der BSD-Lizenz und kann frei verwendet, verändert und weitergegeben werden. Das Moodle-Plugin wurde von seinem zuständigen Entwickler im November 2014 ebenfalls auf diese Version aktualisiert.

Levomepromazin ist ein Antipsychotikum mit Sedationswirkung.
Erstellen Sie zunächst die Strukturformel.



The image shows a screenshot of the JSME (JavaScript Molecule Editor) interface. At the top, there is a text prompt: "Levomepromazin ist ein Antipsychotikum mit Sedationswirkung. Erstellen Sie zunächst die Strukturformel." Below the text is a toolbar with various icons for drawing molecules, including a smiley face, a box, a red X, a red X with a slash, a slash, a curved arrow, a circular arrow, a purple X, and a blue 'I' icon. Below the toolbar is a vertical element palette with buttons for C, N, O, S, F, Cl, Br, I, P, and X. To the right of the palette, the chemical structure of Levomepromazine is displayed. The structure features a benzothiazine ring system with a methoxy group (-OCH3) on the benzene ring and a propylamine side chain (-CH2-CH2-N(CH3)2) attached to the nitrogen atom of the thiazine ring.

Abbildung 3: JSME-Aufgabe

Die Speicherung der Eingaben erfolgt gemäß der Simplified Molecular Input Line Entry Specification (SMILES) oder als Molfile. Eine automatische Auswertung ist über den Zeichenkettenvergleich mit einem Mustermolekül in SMILES-Notation

möglich. SMILES hat jedoch einen entscheidenden Nachteil. Sie ist für Moleküle mit Stereoeigenschaften nicht eindeutig. Daher ist für manche Moleküle eine manuelle Korrektur noch immer nötig. Es gibt einige Alternativen zum JSME, die ein Speicherformat ohne diese Schwäche verwenden. Darunter beispielsweise die Editoren *PubChem* und *ChemDoodle*, die den International Chemical Identifier (InChI) als Speicherformat nutzen. Die Alternativen zum JSME werden getestet und wenn sinnvoll in den Aufgabentyp als Alternative eingebunden.

Der Moleküleditorfragetyp wurde bereits mit Erfolg in Klausuren der MLU eingesetzt und fand gemäß den Klausurevaluationen Akzeptanz unter den Studierenden. Der JSME erfordert eine kurze Einarbeitung für die Prüflinge. An der MLU wurde dies mittels Probeklausuren und Übungsaufgaben ermöglicht. Der Editor erreichte bei einer Softwareevaluation mittels System Usability Scale (SUS) nach einer Klausur mit 51 Teilnehmern den Wert 63, was ihn in puncto Bedienbarkeit im Mittelfeld verortet.

Durch die Zeichenaufgabe sind bildbasierte, interaktive Aufgaben auch in E-Klausuren umsetzbar – wenn auch ohne automatisierte Auswertung. Der Moleküleditor bietet eine Gesamtlösung für einen speziellen Anwendungszweck. Beide Plugins sind auf ilias.de im Forum für Plugins zu finden. Es fehlt eine Zwischenlösung, die einerseits mehrere Freiheitsgrade, aber auch eine sichere automatisierte Auswertung bieten kann.

Key-Feature Fragetyp für Bilder

Die Entwicklung eines neuen Fragetyps wurde im Februar 2015 für die Medizinische Fakultät der MLU begonnen. Prüflinge sollen auf einem Bild (z. B. einer Röntgenaufnahme) Stellen ihrer Wahl markieren und direkt im Bild beschriften. Diese Punkte sind per Drag & Drop verschiebbar. Je nach Beschriftung kann ein Hinweis erscheinen, welcher Bereich exakt zu markieren oder welches Objekt als nächstes zu lokalisieren ist. Einmal gesetzte Punkte können optional nicht mehr verschoben oder umbenannt werden. Eine Punktevergabe erfolgt nach drei Kriterien: Einerseits wird die Beschriftung der Punkte bewertet, wobei eine vom Fragenersteller definierbare Levenshtein-Distanz für den Abgleich mit der Musterlösung berücksichtigt wird. Anschließend wird die Platzierung in Abhängigkeit von der Markierungsbeschriftung bewertet. Diese Bewertung kann auch negativ sein, um etwa patientenschädigende Lösungsversuche besonders zu berücksichtigen. Zuletzt kann optional die Reihenfolge der Markierungserstellung in die Bewertung einbezogen werden.

Die verwendete Technik für diesen Fragetyp basiert auf der quelloffenen Mapping-Software *OpenLayers*. Damit können nicht nur gängige Kartendienste eingebunden werden, sondern auch hochauflösende Bilder. Nutzer können in mit *OpenLayers*

dargestellte Inhalte beliebig hineinzoomen, den betrachteten Bildausschnitt per Drag & Drop verschieben oder in den Vollbildmodus schalten. Auf dem Bild sind zudem mit wenigen Klicks Punkte (durch den Prüfling) und Polygone (durch den Fragenersteller) platziert. Abbildung 4 zeigt ein NASA-Bild des Sirius-Systems, in dem einige Musterlösungsbereiche erstellt wurden. Pro Bereich können unter dem Bildbereich verschiedene Detaileinstellungen getroffen werden.

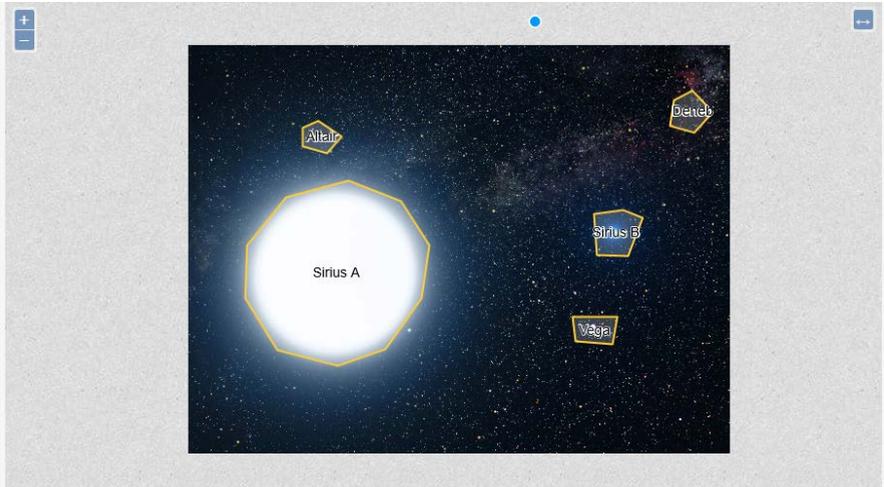


Abbildung 4: Lösungsbereiche inkl. Musterbeschriftung in OpenLayers

Beim von verschiedenen Prüfungsplattformen bekannten Hotspot- / Imagemap-Aufgabentyp sind die potenziellen Lösungsbereiche durch eine Berührung des Mauszeigers erkennbar (er ändert sich zu )¹). Der neue Fragentyp benötigt die Koordinaten der Lösungsbereiche nicht im Prüflingsbrowser. Das Aufgabendesign macht keine Vorgaben, wo etwas zu markieren und wie es zu beschriften ist.

Eingesetzt wurde er in der Medizinischen Fakultät bei einer OSCE-Rotationsprüfung mit 240 Teilnehmern. Es waren dabei an zwei Stationen Röntgenbilder auszuwerten, wichtige Bereiche zu markieren und Anomalien zu finden. Eine Station leitete ein ärztlicher Prüfer, die andere wurde mit ILIAS, diesem Aufgabentyp und der Betreuung einer studentischen Hilfskraft durchlaufen. Die Aufgaben der zwei Stationen wurden nach der Hälfte der Prüflinge getauscht, um einen Bewertungsvergleich durchführen zu können.

Der Aufgabentyp ist noch in der Entwicklung und wird einige kosmetische Überarbeitungen erfahren. Auch die Anforderungspalette an einen Key-Feature Fragentyp

für Bilder ist bislang nicht komplett integriert. Das Zwischenergebnis fungiert als Bildbeschriftungstool mit beschriftungsabhängigen Hinweistexten und automatischer Auswertung. Bereits in dieser Form gibt es zahlreiche neue Möglichkeiten zur Fragenkonzeption. Da OpenLayers ursprünglich eine Kartensoftware mit damit einhergehenden Spezialfunktionen ist, beispielsweise Entfernungs- und Flächenvermessung oder Koordinatenbestimmung, ist insbesondere für die Geowissenschaften Potenzial vorhanden. Der Aufgabentyp wird bei Fertigstellung zumindest eines Großteils der gewünschten Funktionalitäten und nach erfolgreichen Probeeinsätzen der Zeichenaufgabe und dem Moleküleditor auf ilias.de im Forum für Plugins nachfolgen. Die Möglichkeiten von OpenLayers, insbesondere die Markierungsfunktion, können auf dessen Homepage ausprobiert werden (OpenLayers, 2015).

Fazit

Durch anwendungsorientierte Fragetypen und nutzerorientierte Weiterentwicklungen werden E-Klausuren für mehr Lehrende und auch Studierende attraktiver. Nicht nur können viele der einst in Papierform unproblematischen Aufgaben nun in der elektronischen Prüfungsumgebung wieder verwendet werden, auch die Eingaben sind dank der Unterstützungsfunktionen eines Editors eindeutig und obendrein teilweise automatisch auswertbar.

Neue Webtechnologien und die Wünsche der Lehrenden treiben die Entwicklung neuer Fragetypen und anderer Anpassungen weiter voran. Dabei sollte stets beachtet werden, dass Entwicklungen nicht nur spezielle Anforderungen einzelner Personen erfüllen, sondern zumindest für einen bestimmten Fachbereich, im Idealfall sogar universell, einsetzbar sind. Einige ausstehende Entwicklungswünsche, die je nach verfügbarer Zeit konzipiert und umgesetzt werden, sind:

- Textbasierte Key-Feature Aufgaben
- Textauszeichnung und -korrektur
- Drag & Drop Editor für UML-, Flussdiagramme und Zustandsmaschinen
- Herleitung von mathematischen Formeln und Berücksichtigung von Lösungswegen.

Viele andere Entwicklungen sind bereits umgesetzt. Darunter ein browserbasierter Tastaturlayout-Wechsler für die Translatologie oder auch eine einfache digitale Signatur für Testabgaben. Die pluginuntauglichen Entwicklungen sind nicht auf ilias.de zu finden, werden aber auf Nachfrage gern vom LLZ bereitgestellt.

Referenzen

Ertl, P. (2015). JSME Molecule Editor. Abgerufen unter <http://peter-ertl.com/jsme/> [30.04.2015]

Jobst, C. (2014a). Paint question plugin [Forumseintrag]. Abgerufen unter http://www.ilias.de/docu/goto_docu_frm_2528_2891.html [30.04.2015]

Jobst, C. (2014b). JSME (Moleculeeditor) question plugin [Forumseintrag]. Abgerufen unter http://www.ilias.de/docu/goto_docu_frm_2528_2849.html [30.04.2015]

OpenLayers. (2015). Draw and modify features example. Abgerufen unter <http://openlayers.org/en/v3.4.0/examples/draw-and-modify-features.html> [30.04.2015]

Védrine, J. & Stowell D. (2015). Javascript molecular editor question type for Moodle. Abgerufen unter https://github.com/jmvedrine/moodle-qtype_jme [30.04.2015]

UsabilityNet.org. (2015). System usability scale (SUS). Abgerufen unter <http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> [30.04.2015]

Vita

Christoph Jobst, M. Sc. Wirtschaftsinformatik, betreute bereits während seines Studiums in Leipzig E-Learning und E-Assessment verschiedener Lehrveranstaltungen und wechselte 2013 als wissenschaftlicher Mitarbeiter zum LLZ, wo er die softwaretechnischen Belange der Einführung von E-Assessment- und Vorlesungsaufzeichnungssystem verantwortet und bis dato zusammen mit dem Rechenzentrum deren Betrieb sicherstellt. Zusätzlich werden von ihm neue Webdienste und Funktionen, beispielsweise neue Fragetypen für das LMS ILIAS, entworfen und implementiert. Zurzeit promoviert er zum Thema Qualitätssicherung von Aufgaben und Tests in elektronischen Prüfverfahren.

Marcel Dux, Oliver Tietze
Hochschule für Technik und Wirtschaft Berlin

Mündliche Onlineprüfungen auf Distanz: Eine Diskussion der Vor- und Nachteile dieser Prüfungsform

An der Hochschule für Technik und Wirtschaft Berlin (HTW) können Bachelor- und Masterkolloquien als Videokonferenz bzw. Online-Meeting durchgeführt werden. Durch diese Form der Prüfung entstehen für alle Beteiligten deutliche Vorteile bei der Terminfindung, den Kosten und dem zeitlichen Aufwand für die Durchführung. Doch auch neue didaktische und technische Herausforderungen müssen gemeistert werden: die Spielregeln für die professionelle „virtuelle“ Kommunikation sind im Hochschulbereich noch weitestgehend unbekannt.

In diesem Vortrag¹ wurden die Rahmenbedingungen, Abläufe und Besonderheiten von online-begleiteten Kolloquien vorgestellt. Anhand von Praxisbeispielen wurden die Vor- und Nachteile für alle Beteiligten erläutert. Eine kurze Vorstellung des technischen und räumlichen Settings gab einen praxisnahen Einblick in die technischen Möglichkeiten und vermittelte die wichtigsten Regeln für die erfolgreiche Durchführung eines Kolloquiums.

Den Abschluss bildete eine Darstellung der rechtlichen und organisatorischen Veränderungen in den Studien- und Prüfungsordnungen, die ein Angebot in dieser Form erst möglich gemacht haben.

¹ Dux, M. & Tietze, O. (2015). Mündliche Onlineprüfungen auf Distanz: Eine Diskussion der Vor- und Nachteile dieser Prüfungsform: Vortrag auf der Tagung "GML² 2015: Grundfragen Multimedialen Lehrens und Lernens – E-Examinations: Chances and Challenges." Abgerufen unter <http://www.gml-2015.de/tagungsprogramm/dux/> [30.06.2015]

Ausgerichtet vom Center für Digitale Systeme (CeDiS)
der Freien Universität Berlin



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

SUPPORT
QUALITÄTSPAKT
FÜR DIE LEHRE



Learning
Environments
Online



WAXMANN

ISBN 978-3-8309-3331-1



9 783830 933311