



GI-Edition



Lecture Notes in Informatics

**Christoph Igel, Carsten Ullrich,
Martin Wessner (Hrsg.)**

BILDUNGSRÄUME

**DeLFI 2017
Die 15. e-Learning
Fachtagung Informatik**

der Gesellschaft für Informatik e.V. (GI)

**5. bis 8. September 2017
Chemnitz**

Proceedings



Christoph Igel, Carsten Ullrich & Martin Wessner (Hrsg.)

unter Mitarbeit von Maren Braubach

BILDUNGSRÄUME

**DeLFI 2017 – Die 15. e-Learning
Fachtagung Informatik
der Gesellschaft für Informatik**

**5. bis 8. September 2017
Chemnitz, Deutschland**

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-273

ISBN 978-3-88579-667-1

ISSN 1617-5468

Volume Editors

Prof. Dr. habil. Christoph Igel

Deutsches Forschungszentrum für Künstliche Intelligenz DFKI

Technische Universität Chemnitz

10559 Berlin | 09111 Chemnitz, Deutschland

christoph.igel@dfki.de | christoph.igel@cs.tu-chemnitz.de

Dr. Carsten Ullrich

Deutsches Forschungszentrum für Künstliche Intelligenz DFKI

10559 Berlin, Deutschland

carsten.ullrich@dfki.de

Prof. Dr. Martin Wessner

Hochschule Darmstadt

64807 Dieburg, Deutschland

martin.wessner@h-da.de

Series Editorial Board

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria

(Chairman, mayr@ifit.uni-klu.ac.at)

Dieter Fellner, Technische Universität Darmstadt, Germany

Ulrich Flegel, Infineon, Germany

Ulrich Frank, Universität Duisburg-Essen, Germany

Andreas Thor, HFT Leipzig, Germany

Michael Goedicke, Universität Duisburg-Essen, Germany

Ralf Hofestädt, Universität Bielefeld, Germany

Michael Koch, Universität der Bundeswehr München, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Thomas Roth-Berghofer, University of West London, Great Britain

Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany

Torsten Brinda, Universität Duisburg-Essen, Germany

Ingo Timm, Universität Trier, Germany

Karin Vosseberg, Hochschule Bremerhaven, Germany

Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations

Steffen Hölldobler, Technische Universität Dresden, Germany

Thematics

Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2017

printed by Köllen Druck+Verlag GmbH, Bonn

Editorial

Bildung und Qualifizierung findet in Räumen statt. Ob virtuell, hybrid, physisch, ob in Schule und Hochschule, zu Hause oder in Unternehmen, ob in Städten oder ländlichen Regionen. Die Einbettung von Bildung in Unterschiedlichkeiten des Raums beeinflusst und limitiert – ähnlich der Einbettung von Bildung in die Bedingungen der Zeit – das Lehren und Lernen von Menschen. Technologien, digitale Medien und das Internet tragen heute und künftig dazu bei, diese Grenzen zu verändern, mitunter aufzulösen und neue Möglichkeiten des Lehrens, Lernens und Prüfens zu eröffnen.

Die vorliegenden Proceedings der DeLFI 2017 fokussieren mit dem Titel „Bildungsräume“ die Potenziale, Grenzen und Möglichkeiten von Bildungstechnologien und E-Learning in formalen, informellen und non-formalen Bildungsprozessen. In bislang unveröffentlichten Beiträgen aus Forschung, Entwicklung und Innovation wird der State of the Art in der Informatik und in benachbarten Wissenschaftsdisziplinen wiedergegeben, u.a. zu Themen wie Content Engineering, Content Management, Technologien und Tools, Architekturen und Infrastrukturen, Kooperation und Kollaboration, aber auch Fallstudien oder die Betrachtung technischer, sozialer, pädagogischer und organisatorischer Rahmenbedingungen.

Neben der Hochschule, in der die DeLFI traditionell ihren Schwerpunkt hat, wurden bei der 15. DeLFI-Jahrestagung der Gesellschaft für Informatik auch Bildungsräume wie das schulische Klassenzimmer oder der Produktionsarbeitsplatz in der Industrie 4.0 adressiert. Anwendungsfelder wie Medizin, Gesundheitswesen und Sport erwiesen sich als praxisrelevante Bildungsräume für e-Learning und bildungstechnologische Anwendungen.

Einreichungen zum Call for Papers erfolgten für Workshops (Pre-Conferences), als Langbeiträge auf Basis sehr weit fortgeschrittener oder bereits abgeschlossener Arbeiten, als Kurzbeiträge, die laufende oder sich in Vorbereitung befindliche Projekte umfassen oder als Praxisbeiträge, die Projekte mit einem Schwerpunkt auf praktischen Erfahrungen bei Einsatz und Nutzung von e-Learning an beliebigen Lernorten beschreiben. Die Autorinnen und Autoren der Beiträge verfolgten dabei das Ziel, neue wissenschaftliche Erkenntnisse, Praxiserfahrungen, theoretische Arbeiten, konzeptionelle Überlegungen und empirische Ergebnisse für ein Fachpublikum aufzubereiten und vorzustellen.

Unter dem Vorsitz der beiden Program Chairs, Dr. Carsten Ullrich und Prof. Dr. Martin Wessner, trug das Program Committee mit mehr als 50 Wissenschaftlerinnen und Wissenschaftler aus der Informatik und Künstlichen Intelligenz, den Sozial- und Verhaltenswissenschaften sowie den Informations- und Kommunikationswissenschaften mit einem double-blind Peer-Review-Verfahren dazu bei, dass die für eine wissenschaftliche Konferenz erforderliche Sorgfalt und der Anspruch der Veranstalter der DeLFI 2017 an die bestmögliche Qualität von Inhalt und Form gewährleistet wurde. Ihnen allen gilt mein Dank für ihr außerordentliches Engagement in der Sache und für die kollegiale, von Vertrauen und Wertschätzung geprägte Zusammenarbeit.

65 Beiträge wurden für die DeLFI 2017 eingereicht, 3 Gutachten pro Beitrag erstellt, Klarheit und Gewissheit über Annahme oder Ablehnung brachte gleichwohl erst die Präsenzsitzung des Program Committee am 5. Mai in Berlin. So wurden 15 Lang-, 10 Kurz- und 6

Praxisbeiträge ausgewählt, was einer Quote von 47,7 Prozent entspricht. Im Ergebnis liegen nunmehr für die DeLFI 2017 vor: 3 Pre-Conference Workshops zu den Themen: Learning Analytics, Audience Response Systems und VR/AR-Learning. 11 Sessions mit den Schwerpunkten: Adaptivität und Analytics, Kollaboration und Netzwerke, Game-based Learning, Evaluation und E-Assessments, Digitaler Bildungsraum Praxis sowie Systeme und Technologien. Weiterhin 7 Posters und 8 Demos zu ausgewählten Themen des technologie-unterstützten Lehrens, Lernens und Prüfens.

Die 15. DeLFI der Gesellschaft für Informatik wurde gemeinsam mit der 25. Jahrestagung der Gesellschaft für Medien in der Wissenschaft (GMW) geplant und realisiert. Eine herausfordernde Aufgabe für die Veranstalter der vom 5. bis 8. September in Chemnitz durchgeführten Europäischen Fachkonferenz „Bildungsräume“, das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI) und die Technische Universität Chemnitz. Gleichwohl eine lohnende Sache für alle beteiligten Wissenschaftsdisziplinen und Professionen, erfordert eine gemeinsame Veranstaltung nicht nur den Blick über den Tellerrand, sondern auch den ständigen Dialog und schärft das wechselseitige Verständnis für Historien, Kulturen, Prozesse und Zielstellungen.

Ich danke allen Autorinnen und Autoren für ihre Einreichungen sowie die mit diesen Proceedings nunmehr vorliegenden Beiträge zur DeLFI 2017. Danken möchte ich auch unserem Kollegen Dr. Carsten Ullrich, der weit über das zu erwartende Maß hinaus sich für das Gelingen der Fachkonferenz eingebracht hat. Weiterhin Maren Braubach, die mit unermüdlichem Engagement und dem Blick für das Detail die DeLFI-Proceedings 2017 redaktionell betreute. Und abschließend gilt mein Dank dem Leitungsgremium der Fachgruppe E-Learning der Gesellschaft für Informatik für das entgegengebrachte Vertrauen bei der Übertragung der DeLFI 2017.

Berlin, im Juli 2017

Prof. Dr. habil. Christoph Igel

Technische Universität Chemnitz
Deutsches Forschungszentrum für Künstliche Intelligenz DFKI

Tagungsleitung

Christoph Igel, Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Technische Universität Chemnitz (General Chair)

Carsten Ullrich, Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
(Program Chair)

Martin Wessner, Hochschule Darmstadt (Program Chair)

Programmkomitee

Daniel Bodemer
Universität Duisburg-Essen

Lars Bollen
University of Twente

Andreas Breiter
Universität Bremen

Claudia Bremer
Universität Frankfurt/Main

Torsten Brinda
Universität Duisburg-Essen

Mohamed Amine Chatti
RWTH Aachen

Jörg Desel
FernUniversität in Hagen

Jens Drummer
Sächsisches Bildungsinstitut

Martin Ebner
Technische Universität Graz

Wolfgang Effelsberg
Universität Mannheim

Albrecht Fortenbacher
HTW Berlin

Franka Grünewald
Universität Potsdam

Martin Haag
Hochschule Heilbronn

Jörg Haake
FernUniversität in Hagen

Wolfram Hardt
Technische Universität Chemnitz

Andreas Harrer
Technische Universität Clausthal

Peter Henning
Hochschule Karlsruhe

Gerhard Heyer
Universität Leipzig

Ulrich Hoppe
Universität Duisburg-Essen

Marc Jansen
Hochschule Ruhr-West

Paul-Thomas Kandzia
Duale Hochschule Baden-Württemberg

Reinhard Keil
Universität Paderborn

Michael Kerres
Universität Duisburg-Essen

Andrea Kienle
Fachhochschule Dortmund

Ralf Klamma
RWTH Aachen

Johannes Konert
Beuth Hochschule für Technik Berlin

Bernd Krämer
FernUniversität in Hagen

Hanna Krasnova
Universität Potsdam

Detlef Krömker
Universität Frankfurt/Main

Roland Küstermann
Duale Hochschule Baden-Württemberg

Jörn Loviscach
Fachhochschule Bielefeld

Ulrike Lucke
Universität Potsdam

Nils Malzahn
RIAS

Ivana Marenzi
Forschungszentrum L3S

Christoph Meinel
Hasso-Plattner-Institut

Agathe Merceron
Beuth Hochschule für Technik Berlin

Robert Mertens
Hochschule Weserbergland

Wolfgang Müller
Pädagogische Hochschule Weingarten

Niels Pinkwart
Humboldt-Universität zu Berlin

Hans Pongratz
Technische Universität München

Sabine Rathmayer
Hochschule der Bayerischen Wirtschaft

Christoph Rensing
Technische Universität Darmstadt

Ralf Romeike
Universität Erlangen-Nürnberg

Guido Rößling
Technische Universität Darmstadt

Ulrik Schroeder
RWTH Aachen

Andreas Schwill
Universität Potsdam

Marcus Specht
Open University of the Netherlands

Siegfried Stiehl
Universität Hamburg

Sven Strickroth
Universität Potsdam

Stephan Trahasch
Hochschule Offenburg

Karsten Weicker
HTWK Leipzig

Guido Wirtz
Universität Bamberg

Nicole Wöhrle
Universität Freiburg

Raphael Zender
Universität Potsdam

Organisationsteam

Maren Braubach Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Matthias Pohl Technische Universität Chemnitz

mit Unterstützung während der Konferenz von

Andreas Bertsch Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Michael Dietrich Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Julia Ebert Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Antonia Frenzel Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Christiane Karn Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Arash Kermani Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Milos Kravcik Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Vanessa Kretschmar Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Nadine Nagel Deutsches Forschungszentrum für Künstliche Intelligenz DFKI
Michael Schmucker Deutsches Forschungszentrum für Künstliche Intelligenz DFKI

Inhaltsverzeichnis

Digitale Bildungsräume

Christoph Rensing, Wael Alkathib, Max Zollenkopf, Olaf Aschmann, Stephan Tittel

Ein Assistent zur Konzeption von Lernaufträgen für das Lernen am Produktionsarbeitsplatz 15

Dominic Ellek, Wolfgang Müller, Sandra Rebholz, Stefan König

Mobile Web-Technologien im Sport: Konzeption, Entwicklung und Evaluation eines Trainingsassistenten 27

Christian Kohls, Guido Münster

Hybride Lernräume für Innovationsprozesse 39

Falk Pollok, Ulrik Schroeder, Mohamed Amine Chatti

PalmQA: Ein Frage-Antwort-Ensemble für selbstgesteuertes Lernen 51

Julia Finken, Franziska Marx, Michaela Meyer, Philipp Krieter, Andreas Breiter

Entwicklung und Durchführung computerbasierter Tests zur Messung von Musikkompetenzen 63

Benjamin Otto, Till Massing, Nils Schwinning, Natalie Reckmann, Alexander Blasberg, Sandy Schumann, Christoph Hanck, Michael Goedicke

Evaluation einer Statistiklehrveranstaltung mit dem JACK R-Modul 75

Catrina Grella, Nils Karn, Jan Renz, Christoph Meinel

Schulrechner wandern in die Cloud - was bedeutet das für die unterschiedlichen Stakeholder 87

Agnes Koschmider, Mario Schaarschmidt

A Crowdsourcing-based Learning Approach to activate Active Learning 99

Michael Meng, Thorsten Hey, Daniel Peter, Frauke Mattner, Christoph Igel, Christiane Kugler

Akzeptanz von e-Learning Strukturen in Hygieneweiterbildungen 111

Svenja Noichl, Nadine Bergner, Ulrik Schroeder

Zielgruppengerechte App-Icons für Seniorinnen und Senioren 117

Doris Gabler, Martina Müller-Amthor, Georg Hagel

Motiviert an die strukturierte Testfallermittlung 125

Michael Hielscher, Christof Tschudi, Doreen Prasse, Beat Döbeli Honegger

LearningView - ein digitales Werkzeug zur Unterstützung eines offenen Unterrichts 131

Niklas Kreggenfeld, Christopher Prinz, Carsten Ullrich, Bernd Kuhlenkötter

Vorgehensmodell zur Identifikation, Aufnahme und Aufbereitung von Assistenzprozessen in der Industrie 4.0 137

Evaluation und E-Assessment

Kristin Vogelsang, Ilse Hagerer, Uwe Hoppe, Kirsten Liere-Netheler

Entwicklung einer Evaluation für Blended Learning Konzepte 143

Frauke Kämmerer, Heinrich Söbke, Lukas Hartung

Schnell zu erfassen: Ein Komplexitätsmaß für Mehrfachauswahlfragen 155

Julian Dehne, Axel Wiepke, Ulrike Lucke <i>Evaluierung von eLearning - Ein Kommentar zu "Media will Never Influence Learning"</i>	167
Andreas Thor, Norbert Pengel, Heinz-Werner Wollersheim <i>Digitalisierte Hochschudidaktik: Qualitätssicherung von Prüfungen mit dem E-Assessment-Literacy-Tool Eas.LiT</i>	179
Simon Rupp, René Pawlitzek, Carlo Bach <i>Metriken zur Messung von Lernerfolg im Informatik-Grundlagen-Unterricht</i>	185
Bastian Küppers, Florian Kerber, Ulrike Meyer, Ulrik Schroeder <i>Beyond Lockdown: Towards Reliable e-Assessment</i>	191
Elisabeth Katzlinger-Felhofer, Martin Stabauer <i>Digitale Signierung elektronischer Prüfungen: Organisation und Arbeitsumgebung im Prüfungsraum</i>	197
 Adaptivität und Learning Analytics	
Svenja Neitzel, Christoph Rensing <i>Automatische Sammlung von Aktivitäten Lernender in offenenn Lernumgebungen und deren Nutzen in einer Lerntagebuchanwendung</i>	203
Tommy Kubica, Tenshi Hara, Iris Braun, Felix Kapp, Alexander Schill <i>Geführte Auswahl von IT-basierten Lehr/Lern-Werkzeugen</i>	215
Sven Strickroth, Nils Pinkwart <i>Planung von Schulunterricht: Automatisches Feedback zur Reflexionsanregung über eigene Unterrichtsentwürfe</i>	227
Vlatko Lukarov, Ulrik Schroeder <i>Analytics Prototype for Data Driven Decision Making for Blended Learning Strategies in HEI</i>	239
Laura Wartschinski, Le Nguyen-Think, Nils Pinkwart <i>A Conversational Agent for the Improvement of Human Reasoning Skills</i>	249
Michael Striewe, Matthias Kramer, Michael Goedicke <i>Ein Lückentext-Test zur Beherrschung einer Programmiersprache</i>	261
Oliver Schneider, Thomas Martens, Mathias Bauer, Alexandra Ott-Kroner, Uwe Dick, Michael Dorochevsky <i>SensoMot - Sensorische Erfassung von Motivationsindikatoren zur Steuerung adaptiver Lerninhalte</i>	267
 Game-based Learning	
Isabel Slawik <i>Generalisierbarkeit von Gamification-Ansätzen in eLearning - eine explorative Studie</i>	273
Linda Eckhardt, Andreas Pilak, Manuel Löhr, Peer van Treel, Johannes Rau, Susanne Robra-Bissantz <i>Empirische Untersuchung des EGameFlow eines Serious Games zur Verbesserung des Lernerfolges</i>	285
Matthias Ehlenz, Thiemo Leonhardt, Ulrik Schroeder <i>Spielend leicht Lernspiele entwickeln - Ein Framwork für Multitouch-Lernspiele</i>	297

Uwe Arnold, Heinrich Söbke, Maria Reichelt, Thomas Haupt <i>Simulationsfall Nohra: SimCity als etabliertes Lehrmedium in der universitären Hochschulausbildung</i>	303
--	-----

Kollaboration und Netzwerke

Henrik Bellhäuser, Johannes Konert, René Röpke, Christoph Rensing <i>Eine extravertierte und eine gewissenhafte Person in jeder Lerngruppe! Effekte der Verteilung von Persönlichkeitsmerkmalen auf Zufriedenheit und Lernergebnis</i>	309
--	-----

Zhilin Zheng, Nils Pinkwart <i>Recomposing Small Learning Groups at Scale - A Data- driven Approach and a Simulation Experiment</i>	321
---	-----

Ian G. Kennedy, Paul H. Vossen <i>Teamwork assessment and peerwise scoring: Combining process and product assessment</i>	333
--	-----

Sebastian Hobert, Almut Reiners, Pascal Freier, Matthias Schumann <i>StudiDuell App - Mobiles Lernen mit interaktivem Hörsaalspiel</i>	339
--	-----

Systeme und Technologien

Francois Dubois, Truong-Sinh An, Agathe Merceron <i>Eine Middleware Infrastruktur für das Teilen von Lerndaten und Diensten zwischen Bildungseinrichtungen</i>	345
--	-----

Christian Wolters, Daniel Wessel, Finn Jacobsen, Michael Herczeg, <i>Moving freely while staying on track - Smart Glasses to Support Lectures</i>	357
---	-----

Luis Kaufmann, Tobias Welz, Andreas Thor <i>DIAL - Ein BigBlueButton-basiertes System für interaktive Live-Übertragungen von Vorlesungen</i>	369
--	-----

Poster und Demos

Christian Rapp, Bernadette Dilger, Ralf Schimkat, Rainer Mueller <i>Seamless Learning in Lake Constance Region</i>	375
--	-----

Henning Host, Svenja Jaster, Dominik Baedorf, Maria Boos <i>Das E-Portfolio in der Kölner LehrerInnenbildung: Wie kann durch das elektronische Portfolio in ILIAS die kontinuierliche Selbstreflexion von Lehramtsstudierenden an der Universität zu Köln unterstützt werden?</i>	377
---	-----

David Lohner <i>Modulares Lernen online: ein interaktiver Brückenkurs in das Geophysik-Masterstudium (iBRIDGE)</i>	379
--	-----

David Lohner <i>Vergleichbare Bewertung verschiedener videobasierter Lernmodule am Beispiel der Geophysik</i>	381
---	-----

Xenia V. Jeremias, Christian Rabe, Birgit Sellmer <i>Qualitätsgesicherte E-Assessments im Student Life Cycle</i>	383
--	-----

Peter de Lang, Mohsen Shahriari, Ralf Klamma <i>Einführung in das wissenschaftliche Arbeiten und Publizieren mittels Blended Learning in der informatischen Fachdidaktik</i>	385
Majd Edriss, Truong-Sinh An, Francois Dubois, Agathe Merceron <i>Ein Chatbot für die Erstellung von Übungsaufgaben im E-Learning</i>	387
Christopher Mark Fullarton, Tjark Wilhelm Hoeck, Klaus Quibeldey-Cirkel <i>arsnova.click – App für Live-Quizze im MINT-Unterricht</i>	389
Christian Rapp, Jakob Ott <i>Learning Analytics in Academic Writing Instruction - Opportunities Provided by Thesis Writer (TW)</i>	391
Vlatko Lukarov, Ulrik Schroeder <i>AiX-Analytics: Analytics Tool at RWTH Aachen University</i>	393
Sandro Leuchter <i>Mediengestaltung und -einbettung in Moodle zur Vorbereitung von Learning Analytics auf Modulebene</i>	395
Oliver Samoila <i>optes - Mathematik, Lernprozessbegleitung und wie das mit ePortfolios zusammenpasst</i>	397
Johannes Konert, René Röpke, Henrik Bellhäuser <i>mod_groupformation: Moodle Plugin zur algorithmisch optimierten Lerngruppenbildung</i>	399
David Weiß, Patrick Sacher, Daniel Schiffner <i>Das Web Based Training-Toolkit (WBT-T)</i>	401
Verzeichnis der Autorinnen und Autoren	403
GI-Edition Lecture Notes in Informatics	406

Ein Assistent zur Konzeption von Lernaufträgen für das Lernen am Produktionsarbeitsplatz

Christoph Rensing¹, Wael Alkathib², Max Zollenkopf³, Olaf Aschmann⁴, Stephan Tittel⁵

Abstract: Anlernprozesse am Produktionsarbeitsplatz erfolgen heute oftmals individuell gestaltet durch die Bereitstellung von Materialien und über eine Begleitung durch einen erfahrenen Kollegen. In einem Projekt mit Unternehmen entwickeln und evaluieren wir Methoden zur Flexibilisierung, Vereinheitlichung, und Qualitätssicherung der Anlernprozesse. Digital zur Verfügung gestellte Lerneinheiten, die die Lernenden selbständig und in Begleitung von erfahrenen Kollegen bearbeiten, sollen dabei eine zentrale Funktion übernehmen. Da die Lerneinheiten dazu beitragen sollen, insbesondere auch betriebsspezifisches Wissen und Handlungskompetenzen zu erwerben, müssen sie von erfahrenen Kollegen erstellt werden. Der in diesem Beitrag beschriebene Assistent leitet den erfahrenen Kollegen durch den Prozess der Konzeption einer Lerneinheit. Dazu werden zunächst Lernziele in verschiedenen typischen Kompetenzbereichen definiert. Für die einzelnen Lernziele werden durch den Assistenten geleitet Lernaufträge, bestehend aus Lernaufgaben, Lernmaterialien und Lernerfolgskontrollen angegeben. Zum Abschluss der Konzeption erstellt der Assistent einen vorkonfigurierten und teilweise mit Inhalten gefüllten Kurs in der moodle-Lernumgebung. Der Assistent wurde als moodle-Plugin realisiert.

Keywords: Learning Design, Learning Content Authoring, Berufliches Lernen, Lernen am Arbeitsplatz

1 Motivation

Produktionsprozesse in Unternehmen sind heute von einer sehr hohen Dynamik gekennzeichnet. Neue, insbesondere digitale Technologien verändern Produkte und Prozesse. Damit ändern sich die Tätigkeiten der Beschäftigten und es gibt einen gesteigerten Qualifizierungsbedarf. Während Ausbildungsberufe breit angelegte Qualifikationsprofile bieten, erfolgt eine Spezialisierung im Betrieb. Diese Spezialisierung bezieht sich oftmals auf eine konkrete Produktionsanlage oder sogar einen konkreten Arbeitsplatz innerhalb der Produktion. Die Vermittlung der spezifischen Anlagenkenntnisse und der Erwerb der für den Arbeitsplatz notwendigen Qualifikationen erfolgt in aller Regel in einer Kombination aus einer Begleitung des anzulernenden Mitarbeiters durch einen erfahrenen Mitarbeiter und der Bereitstellung von schriftlich verfügbaren

¹ Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstr. 10, 64283 Darmstadt, christoph.rensing@kom.tu-darmstadt.de

² Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstr. 10, 64283 Darmstadt, wael.alkathib@kom.tu-darmstadt.de

³ Kompera GmbH, Lange Rötterstr. 3, 68167 Mannheim, max.zollenkopf@kompera.de

⁴ Kompera GmbH, Lange Rötterstr. 3, 68167 Mannheim, olaf.aschmann@kompera.de

⁵ htcc e.V., Rundeturmstr. 10, 64283 Darmstadt, stephan.tittel@htcc.de

Arbeitsanweisungen. Damit einher gehen Schwachpunkte in der bisherigen betrieblichen Weiterbildungsstruktur in der Produktion:

- Qualifizierungen erfolgen oftmals nicht systematisch und nicht einheitlich gestaltet, sondern individuell unterschiedlich durch die erfahrenen Mitarbeiter.
- Der Prozesse der Einarbeitung ist häufig sehr auf den erfahrenen Mitarbeiter fokussiert. Er zeigt, macht vor, erklärt und beobachtet Lernfortschritte des neuen Kollegen. Damit kommt dem Lehrenden (bei der Gestaltung des Lernprozesses) eher die aktive Rolle und dem Lernenden eher die passive Rolle zu.
- Es besteht eine hohe Abhängigkeit der Qualifizierung von der Verfügbarkeit der erfahrenen Mitarbeiter. Damit fehlen flexible Möglichkeiten, kurzfristig auf betriebliche Bedarfe zu reagieren und einen Kollegen anzulernen.
- Die primäre Aufgabe der erfahrenen Kollegen ist in der Regel nicht die Qualifizierung von Kollegen. Damit besteht die Gefahr, dass der Einarbeitungsprozess nicht immer in der notwendigen Systematik und Tiefe durchgeführt wird. Ein besonderes Augenmerk wird jedoch auf die Aspekte gelegt, die zur Verrichtung der Tätigkeit im engeren Sinn notwendig sind.

Diese Schwachpunkte sind Ausgangspunkt für unsere Projektarbeit. Durch die Nutzung digitaler Medien zur Gestaltung von produktionsbezogenen Lernaufträgen und eine IT-basierte Verfügbarkeit dieser, soll eine Flexibilisierung und qualitative Optimierung des Lernens in der Produktion erreicht werden. Die erfahrenen Kollegen bleiben als Experten der betrieblichen Spezifika innerhalb dieser Zielsetzung unverzichtbar.

Daher konzipieren, entwickeln und evaluieren wir eine Lösung, die erfahrene Mitarbeiter in die Lage versetzt, Lernaufträge zu explizieren, sich mit anderen erfahrenen Kollegen darüber auszutauschen und diese in digitaler Form in einer Lernplattform den anzulernenden Mitarbeitern zur Verfügung zu stellen. Nur unter Einbezug der erfahrenen Kollegen ist es möglich, dass die Lernaufträge praxisnah und auf die Bedürfnisse sowohl des Unternehmens als auch der Mitarbeiterinnen und Mitarbeiter zugeschnitten sind. Die Lernaufträge bestehen zumeist aus einer Kombination von Aufgaben zur eigenständigen Bearbeitung von Lernmaterialien, von Aufgaben, die praktisch zusammen mit dem erfahrenen Kollegen durchgeführt werden müssen und aus Lernerfolgskontrollen. Durch die geeignete Gestaltung der Lernaufträge und deren digitale Verfügbarkeit soll es dem Lernenden ermöglicht werden, den Lernprozess eigenverantwortlicher mitzugestalten.

In diesem Beitrag beschreiben wir das Konzept und die Implementierung eines digitalen Assistenten zur Konzeption der Lernaufträge, der den erfahrenen Mitarbeiter unterstützen soll. Nach einer Betrachtung verwandter Arbeiten, beschreiben wir in diesem Beitrag in Abschnitt 3 die Ergebnisse unserer - in verschiedenen Unternehmen durchgeführten Analysen - zu typischen Lernaufträgen für das Anlernen in der Produktion. Abschnitt 4 stellt das Konzept und die Implementierung des Assistenten vor.

2 Verwandte Arbeiten

Zentraler Inhalt dieses Beitrags ist ein Assistent zur Konzeption von Lernaufträgen. Verwandte Arbeiten in diesem Zusammenhang sind einerseits Modelle und Werkzeuge zur Konzeption von Lernaufträgen und andererseits Autorenwerkzeuge zur Erstellung von multimedialen Lernmaterialien.

Didaktische Entwurfsmuster beschreiben etablierte und erfolgreich erprobte Praktiken der Lehre [Ko09]. Über die in den Mustern gemachten Angaben zum adressierten Problem und den Rahmenbedingungen, lassen sich vom Lehrenden, seinem Ziel entsprechend, geeignete Entwurfsmuster auswählen. Diese Muster machen Angaben zur Realisierung und können von Lehrenden wiederverwendet werden, wenn Sie ihr Lehrarrangement gestalten. Ein digitales Werkzeug zur Auswahl von Entwurfsmustern stellen Frederich et al. [FNN08] mit dem sogenannten didaktischen Assistenten vor. Anhand von Fragen der Form „Welche Art von Wissen wollen Sie vermitteln?“ oder „Welche Merkmale charakterisieren die Zielgruppe?“ und einem Regelwerk werden Alternativen von didaktischen Entwurfsmustern vorgeschlagen.

Mittels IMS Learning Design lassen sich ebenfalls Vorgehensweisen in der Lehre in der Form „Wer führt in welcher Rolle welche Aktivitäten aus? Welche Methoden und welche Umgebung nutzt er dazu, um welches Ergebnis zu erreichen?“ Beschreiben [GB05]. Im Fokus steht dabei die Modellierung des, häufig kollaborativen, Lernprozesses, nicht aber der Lernmedien. Der modellierte Lernprozess kann dann in einer Lernumgebung automatisiert umgesetzt werden. Auch digitale Werkzeuge für die Beschreibung mittels Learning Design existieren [ABL13].

Zur eigentlichen Entwicklung multimedialer Lernmedien dienen Autorenwerkzeuge. Zu unterscheiden sind dabei professionelle Autorenwerkzeuge, die zumeist von Mediendesignern verwendet werden, WSIWYG-Editoren und Rapid Content Authoring Systeme [BHM02]. In der Forschung liegt ein Schwerpunkt auf Ansätzen, die ein Single Sourcing, d.h. die Generierung verschiedener Repräsentationen von Lerninhalten aus einer Quelle, unterstützen [MLT11]. Diese verwenden in der Regel XML-basierte Modelle der Lerninhalte wie <ML>3 [GLT09], weisen aber dementsprechend oftmals gegenüber professionellen Autorenwerkzeugen Schwächen hinsichtlich der Vielfalt der Gestaltungsmöglichkeiten auf. Ein Web-basiertes Autorensystem, das mittels eines integrierten Repositories die Wiederverwendung von einzelnen Elementen einer Lernressource zum Zwecke der Kostenreduktion, beschreibt [HRS05]. Nicht nur die eigentliche Erstellung der Lernmedien, sondern den gesamten Erstellungsprozess beginnend bei der Planung und Konzeption über die Materialsammlung bis hin zum eigentlichen Erstellung der Lernmedien betrachten [Le07]. Ein spezielles Autorenwerkzeug zur Erstellung von kurzen Videoanweisungen für Mitarbeiter in der Fertigung durch die Mitarbeiter selbst stellen Finken et al. [FFU16] vor.

Unserer Zielsetzung besteht darin, einen Assistenten für die Konzeption von Lernaufträgen zu entwickeln. Er soll die, wie im nächsten Abschnitt erläutert wird, besonderen

Eigenschaften von Lernaufträgen zur Vermittlung des Wissens und zum Kompetenzerwerb in der Produktion berücksichtigen. Didaktische Entwurfsmuster sind dafür zu unspezifisch; IMS Learning Design erscheint zu komplex. Insbesondere sind die formale Rollen- und Methodenbeschreibung von Learning Design nicht notwendig. Autorenwerkzeuge fokussieren auf die Medienproduktion. In der Regel ist ein während der Konzeption entstehendes Drehbuch Grundlage für die Produktion selbst. Dennoch sind verschiedene Elemente der verwandten Arbeiten relevant und finden sich im in Abschnitt 4 vorgestellten Konzeptions-Assistenten wieder.

3 Lernaufträgen für Anlernprozesse in der Produktion

Um einen konkreten Einblick in die Herausforderungen und aktuellen Ansätze der Qualifizierung in der Produktion und für den Produktionsarbeitsplatz zu gewinnen, wurden Gespräche mit fünf Unternehmen aus der Chemiebranche geführt. Diese dienten der Analyse der grundsätzlichen Anforderungen an die im Rahmen des Projektes zu entwickelnden IT-gestützten Lösungen. Zusätzlich wurden bestehende Ansätze zur Strukturierung und Unterstützung der Anlernprozesse in den Unternehmen betrachtet. Dazu zählen Kompetenzmatrizen, in denen die verschiedenen, von den Mitarbeitern erworbenen und nachgewiesenen Kenntnisse und Fertigkeiten, deren Erwerb Voraussetzung für den Einsatz an einem Arbeitsplatz sind, dokumentiert werden. Aber auch schriftliche mit Abbildungen angereicherte Arbeitsanweisungen dienen in jedem der betrachteten Unternehmen zur Qualifizierung in Anlernprozessen. Über diese Analyse wurde ein erstes Inventar an Kenntnissen und Fertigkeiten gesammelt. Mittels dieser wurden sechs verschiedene Bereiche von Kenntnissen und Fertigkeiten identifiziert.

Mit einem ausgewählten Unternehmen erfolgte dann schrittweise in mehreren Workshops eine Verfeinerung und Überprüfung dieser Bereiche und die Bestimmung weiterer Elemente der Lernaufträge in Anlernprozessen in der Produktion. Dazu erhielt ein erfahrener Mitarbeiter, der seit vielen Jahren neue Mitarbeiter anlernt, den Auftrag, an konkreten Beispielen zu beschreiben, welches Wissen er vermittelt, welche Kenntnisse und welche Fertigkeiten erworben werden müssen und wie Kenntnisse und Fertigkeiten überprüft werden (können). Anhand mehrerer Beispiele erfolgte dann die Abstraktion vom konkreten Beispiel. Diese Abstraktion wurde dann nochmals überprüft, indem die Abstraktionsebenen als Fragen repräsentiert wurden. Diese Fragen leiteten den erfahrenen Mitarbeiter durch die Beschreibung eines weiteren konkreten Lernauftrags.

Ergebnis dieses iterativen Vorgehens sind die in der Abbildung 1 dargestellten Kenntnisbereiche und Fertigkeiten. Kenntnisse und Fertigkeiten aus diesen Bereichen sind für den Einsatz an jedem Arbeitsplatz in der Produktion zu erwerben. In der bisherigen Vermittlungspraxis stehen die Fertigkeiten zum Ausführen von Tätigkeiten, also die Arbeit an einer konkreten Produktionsanlage, häufig im Fokus der Qualifizierung. Für erfahrenere Mitarbeiter einer Schicht gilt das genauso für den Umgang mit



Abbildung 1: Bereiche von notwendigen Kenntnisse und Fertigkeiten

Störungen. Die Bereiche Umwelt, Sicherheit und Gesundheit und Qualität sind zudem Gegenstand von Qualifizierungsmaßnahmen, zu denen die Unternehmen rechtlich verpflichtet sind. Die Bereiche Prozessverständnis und Grundlagenwissen werden von den anlernenden Kollegen oftmals weniger intensiv betrachtet.

Die von uns ermittelten Bereiche sind ähnlich zu den vier Kompetenzstufen (Orientierungs- und Überblickswissen/ Zusammenhangswissen/Detail- und Funktionswissen/Fachsystematisches Vertiefungswissen) des entwicklungslogischen Lernens in [GSS09]. Sie sind aber spezifischer für den Produktionsarbeitsplatz und um weitere Bereiche ergänzt.

Neben der Bestimmung dieser Bereiche ergab die Analyse ein Strukturmodell, vgl. Abbildung 2, für den Erwerb der für einen Arbeitsvorgang notwendigen Kenntnisse und Fertigkeiten. Ein Arbeitsvorgang umfasst zusammengehörige Tätigkeiten, die ein Mitarbeiter an einem Arbeitsplatz übernimmt. Oberstes Strukturelement des Modells ist ein *Kurs*. Ein Kurs dient der Vermittlung der für einen Arbeitsvorgang notwendigen Kenntnisse und Fähigkeiten. Ein Kurs verfolgt ein *Kurslernziel* (Groblernziel). Ein Kurs kann von einem Mitarbeiter selbständig, ggf. unter Einbezug von Kollegen und Vorgesetzten, bearbeitet werden. Ein Kurs kann ein oder mehrere *Lernaufträge* umfassen. Ein Lernauftrag besteht aus einem *Feinlernziel*, einer oder mehrerer *Lernaufgaben*, *Lernmaterialien* und oft auch *Lernerfolgskontrollen*. Lernaufgaben sind Aufforderungen für den Mitarbeiter, sich mit dem Lerngegenstand, also den zu erwerbenden Kenntnissen oder Fertigkeiten, zu beschäftigen. Die Lernaufgaben sind handlungsorientierte Aufgabenstellungen, deren Durchführung zum Erreichen des Feinlernzieles führen. Eine Lernaufgabe kann im Selbststudium bearbeitet werden, wozu dann Lernmaterialien benötigt werden. Eine Lernaufgabe kann aber auch in Interaktion mit anderen, z.B. erfahrenen Kollegen, bearbeitet werden. Eine Lernaufgabe enthält Übungen, um die erlernten Kenntnisse und Fertigkeiten zu verstetigen.

Mittels einer Lernerfolgskontrolle wird überprüft, ob Feinlernziele erreicht wurden. Es kann sich um eine Selbstkontrolle handeln, mit der der Mitarbeiter für sich selbst überprüft, ob er die Lernziele erreicht hat. Es kann sich aber auch um eine Fremdkontrolle handeln, mit der einem Dritten nachgewiesen werden soll, dass die Lernziele erreicht wurden. Die Lernziele beschreiben in der Regel die zu erwerbenden Kenntnisse und Fertigkeiten. In einem Kurs und dem zugehörigen Groblernzielen sind diese häufig mehrere Kenntnisse und Fertigkeiten. Diese lassen sich den Feinlernzielen der zum Kurs gehörenden Lernaufträge zuordnen.

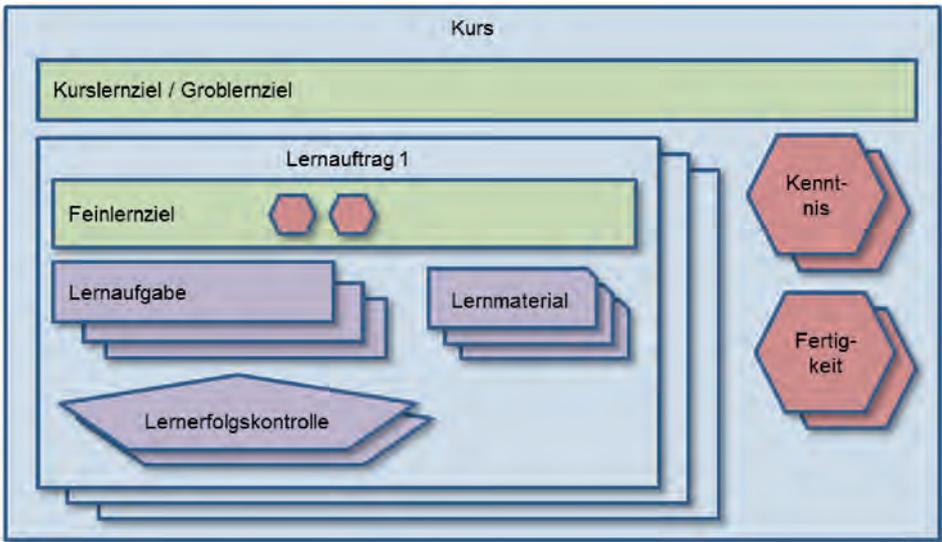


Abbildung 2: Strukturmodell für den Erwerb von Kenntnissen und Fertigkeiten

4 Konzept und Integration des Assistenten in die Lernumgebung

4.1 Konzept

Das aus der Analyse abgeleitete Strukturmodell und die erkannten Bereiche der notwendigen Fertigkeiten und Kenntnisse bilden die Grundlage für das dem Assistenten zugrunde liegende Konzept. In fünf aufeinanderfolgenden Schritten, vgl. Abbildung 3, leitet der Assistent den erfahrenen Kollegen als Autor durch den Prozess.



Abbildung 3: Prozessschritte im Assistenten

Ausgehend von der Beschreibung eines Lernziels für den gesamten Kurs fragt der Assistent den erfahrenen Mitarbeiter in seiner Rolle als Autor nach den Anforderungen an die Tätigkeit. Dabei erfolgt ein Verweis auf die entsprechenden Bereiche, damit der Autor keine Aspekte vergisst. Anschließend definiert der Autor Feinlernziele und ordnet diesen die zuvor als notwendig definierten Kenntnisse und Fertigkeiten zu. Für jedes Feinlernziel erfolgt dann die Feinkonzeption von Lernaufträgen, in welcher Lernaufgaben, Lernmaterialien und Lernerfolgskontrollen benannt werden können.

4.2 Integration in die Lernumgebung

Der Assistent ist als Komponente der Lernumgebung moodle realisiert. Abbildung 4 zeigt die Elemente der Lernumgebung und den Workflow aus den drei Schritten Konzeption des Kurses, Verfeinerung des Kurses und Zugriff auf den Kurs.

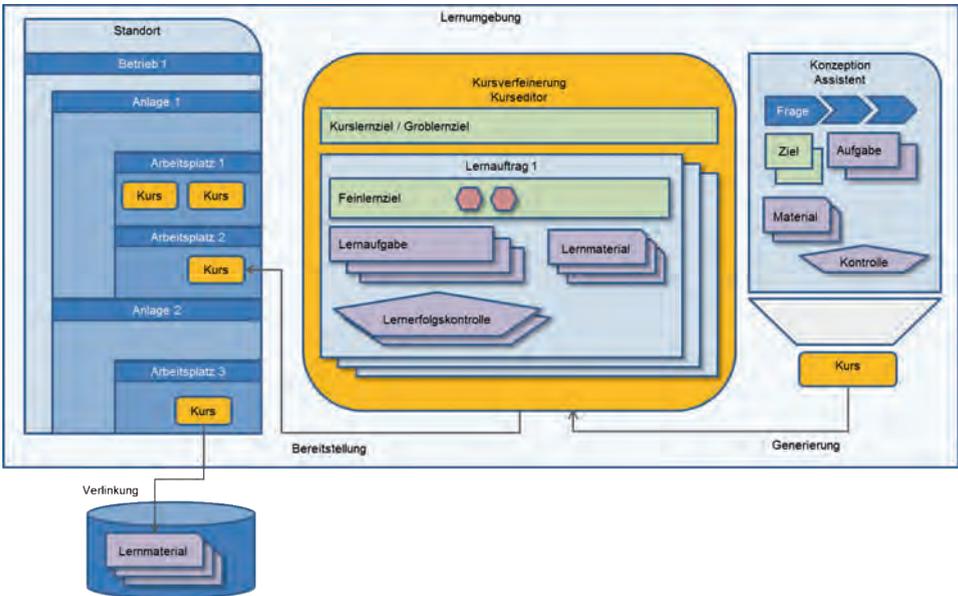


Abbildung 4: Integration des Konzeptionsassistenten in die Lernumgebung

Rechts zu sehen ist der Assistent, mittels dessen Hilfe die Konzeption des Kurses erfolgt. Die Informationen, die bei der Konzeption erfasst wurden, werden beim Abschluss des Assistenten in einen neu generierten Kurs übertragen. Dazu zählen die Benennung des Kurses, dessen Struktur, die Beschreibung von Lernzielen, die Benennung bzw. Beschreibung von Lernmaterial und von Lernerfolgskontrollen. Die eigentliche Ausarbeitung des Kurses erfolgt mittels des Kurseeditors der Lernumgebung. Dieser bietet verschiedenste Werkzeuge zur Gestaltung der Kurselemente an. Über die Abbildung der betrieblichen Struktur erfolgt durch die Beschäftigten der Zugriff auf die einzelnen Kurse.

5 Implementierung

Der Assistent wurde als Plugin für die Lernplattform moodle implementiert. Er leitet den erfahrenen Mitarbeiter entsprechen des Konzeptes des Assistenten in fünf Schritten durch die Konzeption des Kurses. Damit die Mitarbeiter den Kurs später geeignet in der Lernumgebung auffinden können, wird er zunächst der betrieblichen Struktur zugeordnet. Es handelt sich dabei um eine Hierarchie. Im Falle des betrachteten Unternehmens ist sie 3-stufig und unterscheidet Standorte, Betriebe an einem Standort und Anlagen in einem Betrieb. Diese 3-stufige Hierarchie ist in moodle als Hierarchie von Kursbereichen angelegt. Dementsprechend ordnet der Mitarbeiter de-facto den Kurs einem Kursbereich in moodle zu.

The screenshot shows a Moodle course creation form titled "Kurs anlegen". At the top, a green banner contains the text "Im ersten Schritt erfolgt die Anlage eines neuen Kurses." and a "Hilfe" button. Below this, the form is divided into three main sections:

- Kursbereich:** A dropdown menu is set to "Standort Hannover / Betrieb 32 / KS Anlage F21". Below it, a small text block explains: "Die Zuordnung eines Kurses erfolgt nach der betrieblichen Struktur. Ordnungskriterien können sein: Standort, Betrieb, Anlage, Arbeitsplatz. Einen neuen Struktureintrag können Sie [hier anlegen](#)."
- Vollständiger Kursname:** A text input field contains "Wabenkatalysator F21 einstellen und Störungen beheben". Below it, a label reads "Der Name des Kurses."
- Kurzer Kursname:** A text input field contains "Wabenkatalysator F21". Below it, a label reads "Die Kurzbezeichnung wird im Navigationsmenü verwendet."

At the bottom of the form, there are two buttons: "< Zurück" on the left and "Weiter >" on the right.

Abbildung 5: Zuordnung des Kurses und Benennung im Assistenten

Im nächsten Schritt wird der Mitarbeiter aufgefordert, das Lernziel für den Kurs zu benennen. Nachfolgend werden die für die Tätigkeit an einem Arbeitsplatz zu erlernenden Kenntnisse und Fähigkeiten abgefragt. Diese Abfrage erfolgt zunächst als Sammlung. Der Autor erhält über eine Anzeige der verschiedenen Bereiche und zugehörige Leitfragen Unterstützung, wie in Abbildung 6 gezeigt. Im nächsten Schritt erfolgen dann die Benennung der Feinlernziele und die Zuordnung der Kenntnisse und Fähigkeiten zu den einzelnen Lernzielen. Dabei werden die jeweils nicht zugeordneten Kenntnisse und Fertigkeiten im unteren Bereich des Assistenten angezeigt, vgl. Abbildung 7. So soll sichergestellt werden, dass alle Fertigkeiten und Kenntnisse abgedeckt werden.

KEAP Assistent

Start Kurs anlegen Kursziel **Anforderungsprofil** Feinlernziele Lernaufträge

Anforderungsprofil

Im dritten Schritt erfolgt die Erfassung von Kenntnissen und Fertigkeiten, die benötigt werden, um das Kursziel **"Der Mitarbeiter beherrscht die Schmelzeinheit in Anlage B112 zur Einstellung Auftrags an der Anlage."** zu erreichen. [Hilfe](#)

An dieser Stelle werden Kenntnisse und Fertigkeiten gesammelt, um den Arbeitsvorgang souverän ausführen zu können.

Zur Unterstützung finden Sie in der Grafik Leitfragen zu verschiedenen Kompetenzfeldern.

Benennen Sie welche Kenntnisse und Fertigkeiten relevant für den Arbeitsvorgang sind:

Was muss der Mitarbeiter wissen?

1. **Kenntnis:**

2. **Kenntnis:**

Was muss der Mitarbeiter können?

1. **Fertigkeit:**

Leitfrage zu Prozessverständnis:

„Welche relevante Kenntnisse und Fertigkeiten ergeben sich aus den Anlagen- und Arbeitsprozessen?“ [Hilfe](#)

Abbildung 6: Festlegung von notwendigen Fertigkeiten und Kenntnissen

Im letzten Schritt erfolgt dann die Formulierung der Lernaufträge selbst, d.h. deren Feinkonzeption. Je Feinlernziel werden im Assistenten die zugeordneten Kenntnisse und Fertigkeiten angezeigt und dann Materialien, Lernaufgaben und Lernerfolgskontrollen durch den Autoren benannt, wie in Abbildung 8 zu sehen ist. Bei den Materialien ist jeweils anzugeben, um welchen Typ es sich handelt, z.B. ein Dokument, eine Übung oder ein (verlinktes) Onlinetraining. Bei der Lernerfolgskontrolle sind ebenfalls ergänzende Informationen anzugeben, beispielsweise ob es sich um eine Selbst- oder Fremdkontrolle handelt oder ob die Erfolgskontrolle online in der Lernumgebung oder in Präsenz durchgeführt wird.

1. Feinlernziel

Der Mitarbeiter kann den Wabenkatalysator für einen neuen Auftrag einstellen

Zu erlernende Kenntnisse: Zu erlernende Fertigkeiten:

Der Mitarbeiter kann die Aufgabe des Wabenkatalysators im Prozess bes. Der Mitarbeiter kann die Einstellungen des Wabenkatalysators selbstständig vornehmen.

Kenntnis zuordnen...

+ Feinlernziel hinzufügen

Noch nicht zugeordnete Kenntnisse: Noch nicht zugeordnete Fertigkeiten:

Der Mitarbeiter kann die einzelnen Anlagenteile benennen Keine

< Zurück Weiter >

Abbildung 7: Zuordnung der Kenntnisse und Fertigkeiten zu Feinlernzielen

Lernaufträge

Im fünften Schritt erfolgt die vorbereitende Konzeption von Lernaufträgen.

Hilfe

Formulieren Sie für jedes Feinlernziel eine oder mehrere Lernaufgaben. Gestalten Sie die Lernaufgaben so, dass der Lernende durch die Bearbeitung die dem Feinlernziel zugeordneten Kenntnisse und Fertigkeiten erwirbt. Zur Formulierung der Lernaufgaben stehen Textbausteine als Vorlage zur Verfügung. Sie können auch Lernmaterialien und Lernerfolgskontrollen anlegen und den Aufgaben zuordnen.

1. Feinlernziel: Der Mitarbeiter kann den Wabenkatalysator für einen neuen Auftrag einstellen

Zu erlernende Kenntnisse: Der Mitarbeiter kann die Aufgabe des Wabenkatalysators im Prozess beschreiben.

Zu erlernende Fertigkeiten: Der Mitarbeiter kann die Einstellungen des Wabenkatalysators selbstständig vornehmen.

Lernaufgaben Hilfe

Welche Lernaufgaben werden benötigt, um das Feinlernziel zu erreichen?

Lernaufgabe:

Formulieren Sie die Lernaufgabe oder wählen Sie aus den bestehenden Vorlagen aus...

Vorlage Material

+ Lernaufgabe hinzufügen

Lernerfolgskontrolle Hilfe

Wie erfolgt die Lernerfolgskontrolle für das Feinlernziel?

+ Lernerfolgskontrolle hinzufügen

Abbildung 8: Konzeption von Lernaufträgen

Mit dem Abschluss des Assistenten wird dann in moodle ein neuer Kurs erzeugt, dessen Struktur dem Konzept entspricht, welches mit dem Assistenten erstellt wurde. In diesem Kurs werden automatisch einzelne Kursbereiche angelegt und in diese Kursbereiche Informationen eingetragen, die im Assistenten erfasst wurden. Die Materialien, Aufgaben und Lernerfolgskontrollen werden in Form der entsprechenden moodle Materialien oder Aktivitäten ergänzt. Die manuelle Ausarbeitung des Kurses erfolgt dann mittels des moodle Kurseditors und den bestehenden moodle Plugins.

6 Erste Erfahrungen, Zusammenfassung und Ausblick

Der entwickelte Assistent wurde in einem Workshop sechs erfahrenen Mitarbeitern, die an der Analysephase nicht beteiligt waren, vorgestellt. Die Rückmeldungen zum Grundkonzept waren sehr positiv. Der Mitarbeiter sehen in dem Assistenten eine gute Unterstützungsmöglichkeit bei der Konzeption von Lerneinheiten (Kursen), insbesondere für solche Mitarbeiter, die zwar erfahren in der Produktion, aber weniger erfahren in Anlernprozessen sind.

Für die Konzeption von Lerneinheiten existieren bisher nur wenige allgemeine Ansätze zur Unterstützung. Der Schwerpunkt der bestehenden Systeme liegt auf dem eigentlichen Authoring auf Basis der Konzeption. In diesem Beitrag haben wir mit dem Assistenten zur Konzeption von Lerneinheiten diese Lücke geschlossen und dabei die besonderen Aspekte von Lerneinheiten für Anlernprozess in der Produktion berücksichtigt. Der Assistent soll erfahrene Mitarbeiter dabei unterstützen, selbstständig Lerneinheiten zu konzipieren. Erste Rückmeldungen der Anwender sind positiv. Im nächsten Schritt ist vorgesehen, dass mehrere erfahrene Mitarbeiter unter Verwendung des Assistenten Kurse konzipieren. Wir werden diese Nutzungsphase qualitativ evaluieren. Die Ausgestaltung der Lerneinheiten erfolgt dann mittels bestehender Kurseditoren. Nachfolgende Projektaufgaben werden zum einen sein, zu untersuchen ob und wie der Autorenprozess ebenfalls unterstützt werden muss, und zum anderen, ob die konzipierten und erstellten Lerneinheiten für die Qualifizierung der neuen Mitarbeiter in der Produktion geeignet sind.

Danksagung

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PD15010 und des Europäischen Sozialfonds der Europäischen Union (ESF) gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literaturverzeichnis

- [ABL13] Arpetti, A.; Baranauskas, M.; Leo, T.: Making Design Easy: a Usability Evaluation of Latest Generation Learning Design Tools. In World Conference on Educational Multimedia, Hypermedia and Telecommunications, S. 960–965, 2013.
- [BHM02] Baumgartner, P.; Häfele, P.; Maier-Häfele, K.: Auswahl von Lernplattformen: Marktübersicht, Funktionen, Fachbegriffe. In: E-Learning Praxishandbuch, Innsbruck, Wien, Studienverlag, 2002.
- [FFU16] Finken, G.; Freith, S.; Ullrich, C. Implementierung eines Aufnahmewerkzeugs für die semi-automatische Erstellung von Lernszenarien. In (Zender, Raphael, Hrsg.): Proceedings der Pre-Conference-Workshops der 14. E-Learning Fachtagung Informatik, Vol. 1669, CEUR Workshop Proceedings, S. 195-210, 2016.
- [FNN08] Frederich, H.; Niegemann, H.; Niegemann, L.: E-Learning Designentscheidungen in EXPLAIN. In (Loos, Peter; Zimmermann, Volker; Chikova, Pavlina, Hrsg.): Prozessorientiertes Authoring Management: Methoden, Werkzeuge und Anwendungsbeispiele für die Erstellung von Lerninhalten, Logos Verlag, S. 163-182, 2008.
- [GB05] Griffiths, D.; Blat, J.: The role of teachers in editing and authoring units of learning using IMS Learning Design. Advanced Technology for Learning 2/4, S. 243-251, 2005.
- [GLT09] Gries, V.; Lucke, U.; Tavangarian, D.: Werkzeuge zur Spezialisierung von XML-Sprachen für die vereinfachte, didaktisch unterstützte Erstellung von eLearning-Inhalten. In (Schwill, Andreas; Apostopoulos, Nicolas, Hrsg.): Lernen im digitalen Zeitalter, S.211-222. 2009.
- [GSS09] Grantz, T.; Schulte, S.; Spöttl, G.: Lernen im Arbeitsprozess oder: Wie werden Kernarbeitsprozesse (berufspädagogisch legitimiert) didaktisch aufbereitet? In: bwp@Berufs- und Wirtschaftspädagogik – online 17, S. 1-18. 2009.
- [HRS05] Hoermann, S.; Rensing, C.; Steinmetz, R.: Wiederverwendung von Lernressourcen mittels Authoring by Aggregation im ResourceCenter. In (Haake, Jörg M.; Lucke, Ulrike; Tavangarian, Djamshid, Hrsg.): DeLFI 2005: 3. e-Learning Fachtagung Informatik. Köllen Verlag, S. 153-164, 2005.
- [Ko09] Kohls, C.: E-Learning-Patterns. Nutzen und Hürden des Entwurfsmuster-Ansatzes. In (Apostolopoulos, Nicolas; Hoffmann, Harriet; Mansmann, Veronika; Schwill, Andreas, Hrsg.): E-Learning 2009. Lernen im digitalen Zeitalter. Waxmann, S. 61-72, 2009.
- [Le07] Lehmann, L. et al.: Das Authoring Management System EXPLAIN zur ganzheitlichen Unterstützung des Erstellungsprozesses von Trainingsmedien und WBT. In (Eibl, Christian; Magenheim, Johannes; Schubert, Sigrid; Wessner, Martin, Hrsg.): DeLFI 2007: 5. e-Learning Fachtagung Informatik. Köllen Verlag, S. 139-150, 2007.
- [MLT11] Meincke, F.; Lucke, U.; Tavangarian, D.: Empfehlungen zur Nutzung eines Textverarbeitungswerkzeugs zur Erstellung von XML-basierten E-Learning Inhalten. In (Rohland, Holger; Kienle, Andrea; Friedrich, Steffen, Hrsg.): DeLFI 2011: Die 9. e-Learning Fachtagung Informatik. Köllen Verlag, S. 9-20, 2011.

Mobile Web-Technologien im Sport: Konzeption, Entwicklung und Evaluation eines Trainingsassistenten

Dominic Ellek¹, Wolfgang Müller², Sandra Rebholz³ und Stefan König⁴

Abstract: In diesem Beitrag wird die Konzeption, Entwicklung und Evaluation eines technologiebasierten Trainingsassistenten vorgestellt. Der Prototyp des mobilen Trainingsassistenten mit der Bezeichnung TASI wurde für den Bereich Handball konzipiert. Mit ihm sollen typische Problemfelder bewältigt und alternative Trainings- und Organisationsformen, wie beispielsweise individuelles Training, unterstützt werden. Für die wissenschaftlich fundierte Entwicklung des Prototyps wird der Design-Based Research-Ansatz aufgegriffen. Hinsichtlich dessen findet eine formative und summative Evaluation statt. Der Trainingsassistent wurde einer Testphase im Trainingsbetrieb unterzogen und anhand eines Fragebogens, basierend auf dem Technology Acceptance Model (TAM) und der System Usability Scale (SUS), untersucht. Mit den verschiedenen Erhebungsmethoden konnte beobachtet werden, dass der Assistent von den Nutzern akzeptiert wurde und einen Mehrwert generiert hat.

Keywords: Mobiler Trainingsassistent, Web-Technologien im Sport, Handball, Design Based-Research, Responsive Design, Technology Acceptance Model, System Usability Scale.

1 Einführung

Die Entwicklung im Sport sorgt für einen zunehmenden Gebrauch von digitalen Medien. Mobile Technologien finden heutzutage in der Schule, im Vereins-, Freizeit- oder Profisport Verwendung. Die Potentiale sind dabei längst nicht ausgeschöpft. Vor allem den digitalen online Medien wird eine unterstützende Wirkung im Lernprozess zugeschrieben [HKV13]. Das Medium Video spielt dabei eine wichtige Rolle. Mit den aktuellen Technologien ist beispielsweise auch im Amateurbereich ein mobiles Video-Feedbacktraining möglich [Kr15]. Im Internet, vor allem auf sozialen Plattformen, wird täglich eine Vielzahl von Sport- und Trainingsvideos veröffentlicht. Mobile Endgeräte ermöglichen eine einfache Produktion, Verarbeitung und Bereitstellung. Der Einsatz von digitalen Online-Technologien kann zur Bewältigung typischer Probleme im sportlichen Training führen. Auch liegt es nahe, dass Veränderungen und Innovationen im Training auf diese Weise besser eingeleitet und umgesetzt werden können.

¹ PH Weingarten, Mediendidaktik und Visualisierung, Kirchplatz 2, 88250 Weingarten, tasi.trainingsassistent@gmail.com

² PH Weingarten, Mediendidaktik und Visualisierung, Kirchplatz 2, 88250 Weingarten, mueller@md-phw.de

³ PH Weingarten, Mediendidaktik und Visualisierung, Kirchplatz 2, 88250 Weingarten, rebholz@md-phw.de

⁴ PH Weingarten, Sport, Kirchplatz 2, 88250 Weingarten, koenig@ph-weingarten.de

2 Problemstellung, Ansatz und Zielsetzung

Innovative Ansätze im Sport, wie beispielsweise individuelles Training im Handball, werden im Breitensport kaum umgesetzt, da die hierfür notwendigen Ressourcen in der Regel nicht vorhanden sind. Tatsächlich führen in der Praxis knappe Hallenkapazitäten und die geringe Anzahl an ausgebildeten Trainern zu einer Abnahme der Trainingsqualität. Das Ergebnis einer Befragung von über 500 Lehrern im Rahmen der LEARNTEC 2016 [Le16] verdeutlichte, dass die Akzeptanz gegenüber neuen Medien fehle. Zudem kritisieren Hebbel-Seeger, Kretschmann und Vohle [HKV13] die mangelnde Umsetzung der Erkenntnisse aus den wissenschaftlichen Projekten in diesem Bereich und deren ausbleibende Weiterführung.

In diesem Beitrag wird ein Ansatz auf Grundlage digitaler Medien und speziell mobiler Web-Technologien vorgeschlagen, der [Kr15] folgend im Kontext dieser Problemfelder Lösungsansätze bieten soll. Konkret wird das Konzept eines technologiebasierten Trainingsassistenten vorgestellt und evaluiert, der individualisiertes Training in der Breite ermöglichen und damit im Breitensport bislang in dieser Form nicht vorhandene Trainingsformen zulassen soll, wie etwa zusätzliche Trainingseinheiten oder Unterstützung beim Training am Seitenstreifen. Für die Entwicklung eines solchen technologiebasierten Trainingsassistenten muss zunächst geprüft werden, in welchen Szenarien der Einsatz von digitalen Medien im Handballtraining sinnvoll ist. Insbesondere muss dabei aufgezeigt werden, welchen Mehrwert ein digitaler Trainingsassistent für Trainer und Spieler bieten kann und wie der Trainingsassistent gestaltet sein muss, um von den zukünftigen Nutzern akzeptiert zu werden.

3 Stand der Technik

Die Nutzung von digitalen Medien zur Veranschaulichung einzelner Bewegungen, vor allem im Bereich des Technikerwerbstrainings, gab es schon in den 1970er Jahren [HKV13]. In den frühen 90er Jahren wies Ihlo [Ih91] dem Einsatz von Medien folgende Eigenschaften zu: Motivierung, Stimulierung, Erlebnisvermittlung, Steuerung, erhöhte Effektivität, Kontrolle und Rationalisierung im Sinne von Zeitersparnis in der Umsetzung und Durchführung. Seit dem Einzug der Digitalisierung sind der Videografie kaum noch Grenzen gesetzt. Dreidimensionale Animationen, Bewegungsabläufe, animierte Einzelbilder, Videos mit Zeitlupen- oder Zeitrafferfunktion und Bild-in-Bild Sequenzen sind nur ein kleiner Auszug der aktuellen Technik. Hinzu kommt die einfache Bearbeitungs- und Schnittfunktion von Videomaterial. Die Programme und Applikationen sind oft kostenfrei und einfach zu bedienen. Die Möglichkeit der globalen Veröffentlichung über das Internet und deren Kommunikations- und Kommentarfunktionen sind weitere wichtige Bestandteile.

Den Sinn und Zweck der Nutzung von Medien im Sport sieht die Literatur in der Vermittlung von Bildung. Dabei spielt es keine Rolle, ob es sich um Schul-, Hochschul-, oder Vereinssport handelt. Das Medium kann aber nur bildend wirken, wenn es beim

Nutzer eine Reflexion über die Eigene- und die Fremdbewegungen, das gemeinsame Spiel oder die Taktik veranlasst. Dabei wird vor allem den digitalen Online-Medien eine unterstützende Wirkung im Lernprozess zugeschrieben [HKV13]. Praxisbeispiele wie Videofeedback-Training im Volleyball [FTB05] und Handball [Kr15], Video-Podcasts im Sportstudium zum Thema Snowboarden (Projekt „RUN! – eLearning in Bewegungs- und Sportwissenschaft“; boardcast.de) oder Blended Learning in der Traineraus- und –fortbildung im Tischtennis (edubreak-sportcampus.de) veranschaulichen das.

Allgemein wird der geringe Einsatz von aktueller Bildungstechnologie in der Forschung und im Schulsport bemängelt [HKV13]. Im Freizeit- oder Fitnesssport finden mobile Technologien dagegen immer häufiger Verwendung. Fitness-Apps finden großen Anklang und stehen unter ständiger Weiterentwicklung. Auch Sport-Apps wie *online-coach* (online-coach.org) oder die App *Coach's Eye* (coachseye.com) beinhalten nützliche Funktionen, wie beispielsweise das mobile Abrufen von selbst zusammengestellten Trainingseinheiten oder das Analysieren von selbst aufgenommenen Videos mit Hilfe von Zeitlupenfunktion, Audiokommentaren und Zeichenwerkzeugen. In Bezug zu den bereits erörterten Problemfeldern gibt es, speziell im Bereich Handball, keine erwähnenswerten Medienprodukte. Zwar beinhaltet beispielsweise die kostenpflichtige App *online-coach* auch Handballübungen, doch mangelt es dabei an der sportlich-didaktischen Aufbereitung dieser. Ebenso ist keine spezielle Unterstützung für individuelles Training oder für die Trainingsevaluation vorgesehen.

4 Konkrete Forschungsfragen

Auf Basis des dargestellten Forschungsbedarfs ist das Ziel der Arbeit, einen mobilen Trainingsassistenten für die Sportart Handball zu entwickeln und dessen Akzeptanz bei den Nutzern zu evaluieren. Um den Kontext spezifischer zu untersuchen, werden folgende Forschungsfragen gestellt.

F₁: Wird der mobile Trainingsassistent von den Nutzern in der Anwendung akzeptiert?

F₂: Welchen Mehrwert generiert der mobile Trainingsassistent?

Die Vermutung liegt nahe, dass beide Fragen direkt voneinander abhängen. Würde der Assistent dem Nutzer keinen Mehrwert bringen, so würden ihn diese wohl kaum weiterverwenden (akzeptieren). Diese Forschungsfragen lassen sich durch weitere Unterfragen konkretisieren: *F_{2,1}: Welchen Mehrwert hat der Trainingsassistent für den Trainer im Training?* *F_{2,2}: Welchen Mehrwert generiert der Trainingsassistent für den Sportler?*

5 Methodik: Design-Based Research

Hinsichtlich der Thematik ist es notwendig, einen Ansatz zu wählen, der eine

Verknüpfung zwischen erkenntnis- und anwendungsorientierter Forschung darstellt. Der Design-Based Research (DBR) Ansatz vereint Entwicklung und Forschung in einem kontinuierlichen Kreislauf von Gestaltung, Durchführung, Überprüfung und Re-Design [De03], [Re05]. Mit dem Ansatz möchte man konkrete und realitätsbezogene Lösungen zu relevanten (Bildungs-)Problemen finden. Für den DBR-Ansatz gibt es in der Literatur verschiedene Bezeichnungen. Im Deutschen verwendet Reinmann den Begriff *entwicklungsorientierte Bildungsforschung* bzw. *didaktische Entwicklungsforschung* [Re14]. Der Ansatz ist daher besonders für die Entwicklung von neuartigen mobilen Lehr-Lerntechnologien geeignet [HHM09], und erweist sich somit zur Beantwortung für die vorliegenden Forschungsfragen als angemessen.

6 Konzeption

Mit Hilfe einer Anforderungsanalyse wird die genaue Beschreibung der notwendigen Bestandteile und Funktionen des Trainingsassistenten erstellt. Dabei werden, im Rahmen eines szenariobasierten Designs (SBD) Personas und Problem-Szenarien erzeugt. Die daraus resultierenden Problemfelder werden in einen Anforderungskatalog für den Trainingsassistenten überführt. Unter Anbetracht dieser Anforderungen und den bereits erstellten Forschungsfragen, findet die Hypothesenbildung statt. Im Rahmen des SBD wird dem Trainingsassistenten der Arbeitstitel *TASI* erteilt. Dieser findet fortan Verwendung und individualisiert den Prototyp.

6.1 Problemfelder und Anforderungen

Die nachfolgenden Problemfelder im Bereich Handballtraining stützen sich auf Problemszenarien [CR90], welche auf Grundlage von Personas [Co04] generiert wurden: Das Training ist nicht abwechslungsreich. Die Trainingsplanung und die Nachbereitung ist zeitintensiv. Spieler sind im Training oft unmotiviert. Ein Grund dafür ist die mangelnde Trainingsqualität. Ein individuelles Training ist mit den vorhandenen Ressourcen nur schwer durchzuführen. Die Trainerqualität und damit auch die Qualität der Übungen ist oft unzureichend.

Auf Basis dieser Problemfelder ergeben sich somit folgende Anforderungen an den Trainingsassistenten: Der Nutzer möchte neue Übungen recherchieren und im Trainingsbetrieb verwenden können. Der Assistent soll die Planung und Evaluation des Trainings optimieren und zeitlich verkürzen. Außerdem soll er die Trainingsausführung verbessern, um die Trainingsqualität zu erhöhen. Zusätzlich soll er individuelles Training ermöglichen und fördern. Grundlage für eine erfolgreiche Nutzung ist eine einfache, schnelle und intuitive Bedienbarkeit.

6.2 Definition Mehrwert und Hypothesenbildung

Aus den Ergebnissen der Anforderungsanalyse lässt sich der potentielle Mehrwert des

mobilen Trainingsassistenten für Trainer und Spieler ableiten und folgendermaßen beschreiben:

Der mobile Trainingsassistent unterstützt den Trainer in allen drei Phasen des Trainings, das heißt, er vereinfacht die Organisation in der Planung, Ausführung und Evaluation. Die Vor- und Nachbereitung des Trainings ist mit Hilfe des Trainingsassistenten einfacher und effizienter zu gestalten. Nach einer Eingewöhnungsphase kann man mit TASI zeitsparend arbeiten. Die Nachbereitung des Trainings, welche oft vernachlässigt wird, kann durch TASI halbautomatisch erfolgen. Zudem ermöglicht der Assistent eine Archivierung der Trainingseinheiten. So kann kontrolliert werden, was trainiert wurde. Dies verbessert auch die Trainingssteuerung [Ho96], da man abwechslungsreicher und auf bestimmte Übungen gezielt aufbauend, trainieren kann. Mit einer wachsenden Anzahl an Übungen stellt TASI eine Art Übungsbibliothek dar. Somit hat der Nutzer jederzeit die Möglichkeit Übungen nachzuschauen, selbst auszuüben oder in das Training zu integrieren. Die Trainingsdurchführung kann abwechslungsreicher gestaltet werden, da immer wieder neue Übungen eingesetzt werden können. Hinzu kommt, dass die ausgewählten Übungen die korrekte Umsetzung (Technik) nach neuestem Stand der Trainingslehre beschreiben und somit die Trainingsqualität erhöhen. Das heißt, durch die audiovisuelle Aufbereitung der Bewegungsabläufe kann der Nutzer genau erkennen, wie er die Übung auszuführen hat. Der Trainer wird somit gar nicht erst verleitet eine Übung unpräzise bzw. „falsch“ vorzuführen. Den Spielern wird im Video das ideale Grundmuster des Bewegungsvorgangs präsentiert.

Der Mehrwert wird im Folgenden mit dem Begriff *Nützlichkeit* gleichgesetzt. Die *wahrgenommene Nützlichkeit* (perceived usefulness, kurz: PU) wird von Davis, Bagozzi und Warshaw [DBW89, S. 985] folgendermaßen definiert: „the prospective user’s subjective probability that using a specific application system will increase his or her job performance within an organizational context“. Aus der wahrgenommenen Nützlichkeit kann demnach abgeleitet werden, ob das System aus Sicht des Nutzers einen Mehrwert für ihn bietet. Laut dem Technology Acceptance Model (TAM) nach Davis et al. [DBW89] ist die wahrgenommene Nützlichkeit eine der zwei Variablen, von denen die Verhaltensakzeptanz abhängig ist. Die andere Variable ist die *wahrgenommene einfache Bedienbarkeit* (perceived ease of use, kurz: PEoU) [BM04]. Die Nützlichkeit (also der Mehrwert) und die einfache Bedienbarkeit sind für die tatsächliche Nutzung ausschlaggebend. Bürg und Mandl setzen die tatsächliche Nutzung mit der Verhaltensakzeptanz gleich. Voraussetzung dafür ist eine positive Nutzungseinstellung, welche im Kontext synonym zur Einstellungsakzeptanz verwendet wird [BM04]. Aus der wahrgenommenen Nützlichkeit und der wahrgenommenen einfachen Bedienbarkeit kann demnach abgeleitet werden, ob ein System vom Nutzer akzeptiert wird.

Auf dieser Grundlage lässt sich zusammenfassend die Hypothese formulieren: *H: Der Trainingsassistent wird von den Nutzern akzeptiert, da er einfach zu bedienen ist und einen Mehrwert für den Nutzer mit sich bringt.* Wird die Aussage in ihre Einzelteile zerlegt, können weitere Unterhypothesen gebildet werden: *HU₁: Der Trainingsassistent ist einfach zu Bedienen. HU₂: Der Trainingsassistent hat einen Mehrwert für den Nutzer. HU₃: Der Trainingsassistent wird vom Nutzer akzeptiert.*

7 Umsetzung

Eine erste Umsetzung des mobilen Trainingsassistenten erfolgt unter Verwendung verschiedener Webtechnologien. Da der Trainingsassistent sowohl eine Authoring-Umgebung zum Zusammenstellen von Trainings und deren Zuweisung an Mannschaften und Teams beinhalten soll, als auch eine einfache Umgebung zum Abrufen der Trainings und der dazugehörigen Videos auf dem mobilen Endgerät, wurde eine dynamische Website im Responsive Design als besonders geeignet befunden. Das Framework Material Design for Bootstrap (MDB) unterstützt den Designprozess. Als Skriptsprache wird PHP und als Datenbanksystem MySQL verwendet.

Für Prototyp 1 wurde ein erstes Testsystem mittels XAMPP und dem MDB Framework aufgezogen. So konnte mit HTML5, CSS3 und Javascript ein Grundgerüst und ein funktionsfähiges Klickmodell initiiert werden. Prototyp 1 wurde im Rahmen eines kurzen Usability-Tests mit der Thinking-Aloud-Methode geprüft. Es konnten keine gravierenden Mängel festgestellt werden. Kleinigkeiten in der Darstellung wurden daraufhin optimiert. Bei der Umsetzung von Prototyp 2 wurde der Funktionsumfang minimiert und nur die Strukturen implementiert, welche zur Beantwortung der Forschungsfragen nötig sind. Zudem wurden über 100 Trainingsübungen aus verschiedenen Bereichen ausgewählt, abgefilmt, auf eine Videoplattform geladen und in die Datenbank eingetragen. Dabei wurden folgende Funktionen verwirklicht: Der Assistent passt sich mittels Responsive Webdesign auf die verschiedenen Displaygrößen des Endgeräts an. Die erstellten Übungsvideos lassen sich nach Kategorien anzeigen. Mit Hilfe der Login-Funktion können Trainer diese Übungen zu einem Training hinzufügen, mit Hinweisen versehen und für die Nutzer freigeben (s. Abb. 1).

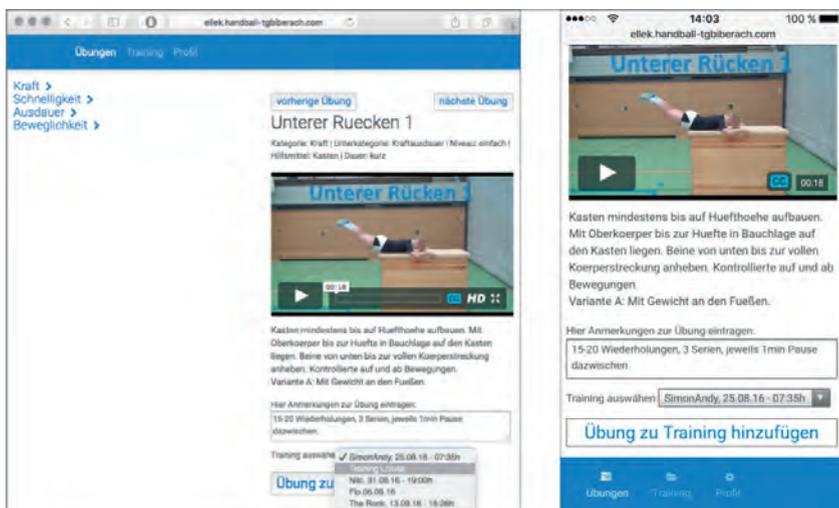


Abb. 1: TASI in Verwendung auf einem Notebook (links) und auf einem Smartphone (rechts)

8 Evaluation

Im Rahmen des DBR-Ansatzes kamen verschiedene Untersuchungsmethoden (Beobachtung, Befragung, Testphase, Feedback, Usability-Testing, Selbstevaluation und Experten-Einschätzung) zum Einsatz. Zu Beginn der Forschungsarbeit wurde formativ evaluiert. Ergebnisse konnten somit direkt in den Entwicklungsprozess zurückfließen. Ebenso fand abschließend eine summative Evaluation statt. Ziel der Evaluation war es, die eingangs formulierten Forschungsfragen zu beantworten und die aufgestellten Hypothesen zu überprüfen. Um herauszufinden, ob die Nutzer den mobilen Trainingsassistenten akzeptieren und welchen Mehrwert sie in der Verwendung des Assistenten sehen, wurde eine schriftliche Befragung aller Teilnehmer, ein Leitfadenterview der Trainer und verschiedener Experten aus dem Bereich Sport, sowie eine Beobachtung des Trainings während der Testphase durchgeführt. Für die Evaluationsplanung wurden zwei Testszenarien entwickelt, welche nachfolgend kurz beschrieben werden.

8.1 Testphase

Testszenario 1 stellt die Testphase mit Trainern und Spielern dar. Vorgabe war es, den Trainingsassistenten TASI im Training einzusetzen (siehe Abb. 2). Die Durchführung fand in zwei Gruppen (Gruppe A und B) statt, da die Probanden zu zwei unterschiedlichen Mannschaften gehörten. Der jeweilige Trainer integrierte TASI in eine oder mehrere Trainingseinheiten. Diesbezüglich kam der Assistent auch in der Trainingsvorbereitung und -nachbereitung zum Einsatz. Im Anschluss an die Testphase fand eine schriftliche Befragung der Trainer und Spieler zur wahrgenommenen Nützlichkeit und zur wahrgenommenen Einfachheit der Bedienung des Trainingsassistenten mit Hilfe eines Fragebogens statt. Zusätzlich wurde ein Leitfadenterview mit den Trainern durchgeführt, das einen Einblick in deren Beurteilung des Systems verschaffen sollte.

Um weitere Stakeholder mit einzubeziehen, wurde Testszenario 2 entwickelt. Dabei wurden drei Experten aus verschiedenen Sportbereichen (Sportwissenschaftler, Athletiktrainer, aktive Handballspielerin) hinsichtlich der Verwendung von TASI einem Nutzungstest samt mündlicher Befragung unterzogen. Die Befragung sollte Hinweise darauf liefern, welchen Mehrwert die Experten in der Nutzung des Trainingsassistenten sehen.

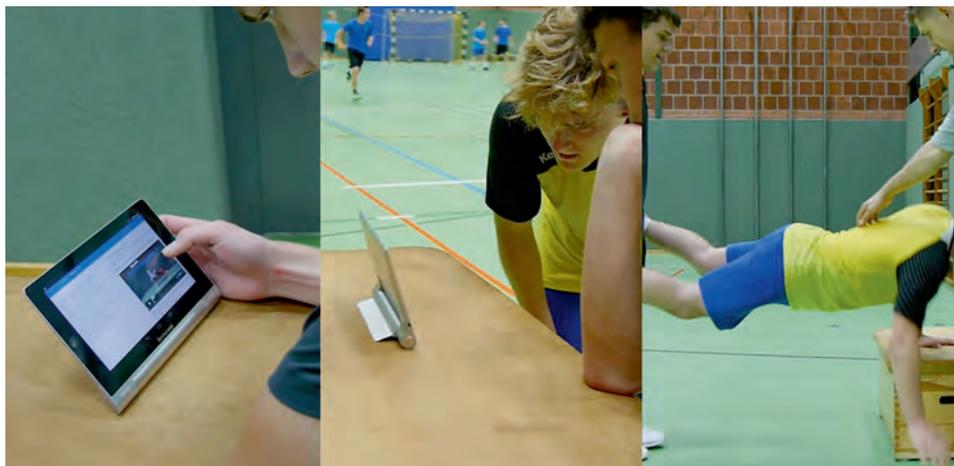


Abb. 2: Training mit TASI am Seitenstreifen in der Halle (Testszenario 1, Gruppe A)

8.2 Auswertung

Alle Probanden der beiden Gruppen aus Testszenario 1 wurden mit Hilfe eines Fragebogens befragt, der aus drei Teilen bestand. Im ersten Teil wurden demographische und persönliche Daten erfasst. *Gruppe A* setzt sich aus 14 männlichen Spielern zusammen und weist ein Durchschnittsalter von 17,64 Jahren auf (SD: 2,373). Die Spieler sind im Durchschnitt seit 10,71 Jahren beim Handball (SD: 3.872). In einer Woche bestreiten sie durchschnittlich 3 Trainingseinheiten Handball (SD: .864). Der Trainingsassistent wurde von der Gruppe vorwiegend auf einem Tablet genutzt (86%). Durchschnittlich 2,5 Mal kam TASI je Proband im Trainingsbetrieb zum Einsatz. *Gruppe B* setzt sich aus 6 weiblichen Spielerinnen zusammen und weist ein Durchschnittsalter von 15,17 Jahren auf (SD: .753). Die Spielerinnen sind im Durchschnitt seit 7,83 Jahren beim Handball (SD: 3.251). In einer Woche bestreiten sie durchschnittlich 4 Trainingseinheiten Handball (SD: .983). Der Trainingsassistent wurde von *Gruppe B* in genau einer Trainingseinheit auf einem Smartphone genutzt.

Der zweite Teil des Fragebogens beinhaltet 12 Items (Item 9-20) die auf Grundlage des bereits erwähnten Technical Acceptance Model (TAM) konzipiert wurden. Der durchschnittliche PEoU-Wert von 4,55 proklamiert einen geringen Aufwand in der Nutzung des Systems. Somit besteht laut Davis et al. [DBW89] eine höhere Wahrscheinlichkeit, dass das System auch tatsächlich genutzt wird. Der durchschnittliche PU-Wert von 4,18 stellt den erwarteten Ertrag dar. Der Nutzer geht in diesem Falls also davon aus, dass ihm durch die Verwendung des Systems ein Mehrwert entsteht. Somit ist er auch eher dazu bereit, das System zu nutzen. In Abbildung 3 werden die genannten Kennzahlen zusätzlich gruppenspezifisch dargestellt. Dabei lässt sich erkennen, dass in *Gruppe B* leicht höhere Werte zu vernehmen sind.

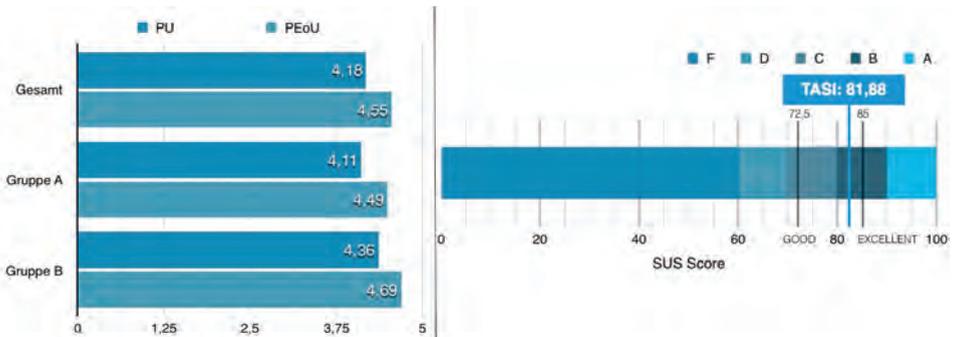


Abb. 3: Auswertung TAM (links) und SUS (rechts)

Im dritten Teil des Fragebogens wird mit 10 Items (Item 21-30) die System Usability Scale (SUS) [Br96] ermittelt. Die Auswertung ergab, wie in Abbildung 3 rechts zu erkennen, einen Wert von 81,88. Somit bestätigt dieses Ergebnis die subjektive Einschätzung der Gebrauchstauglichkeit (PEoU). Dem Trainingsassistenten wird daher eine gute Usability ohne größere Probleme zugewiesen.

Die Trainer beider Gruppen und auch die drei Stakeholder aus Testszenario 2 wurden mittels eines Leitfrageninterviews befragt. Deren Auswertung lässt schlussfolgern, dass die Verwendung von TASI positiv verlaufen ist. Die Trainer beider Gruppen verwendeten den Trainingsassistenten auf ähnliche Art und Weise. Alle Funktionen wurden genutzt und somit auch getestet. Beide Trainer bemängelten das automatische Schließen der Navigation. Insgesamt wurde TASI eine gute Bedienbarkeit zugesprochen. Eine Darstellung der ungefähren Durchführungsdauer und des Mindestalters zur Übungsdurchführung wurden vorgeschlagen. Die Befragten sehen die Verwendung von TASI im individuellen oder Kleingruppen-Bereich sowie im Athletik- und Torwarttraining. Ein Proband aus Testszenario 2 bestätigt die Annahme, dass es vor allem auf den Demonstrator der Übung ankommt, inwiefern der Bewegungsablauf im Video korrekt dargestellt wird. Derselbe Proband bewertet die audiovisuelle Beschreibung, vor allem in Anbetracht der kurzen Dauer der Videos, als positiv. In der geringen Zeit der Videos sei es aber nicht möglich die Übung in vollster Genauigkeit, hinsichtlich des sportwissenschaftlichen Hintergrunds, zu erklären. Ein Proband hat alle 115 Übungsvideos angeschaut. Dabei empfand er das 2-sekündige Intro als störend. Besonders hervorzuheben ist die Tatsache, dass zwei Experten die didaktisch-methodische Umsetzung sowie den sportwissenschaftlichen Inhalt der Übungsvideos durchweg positiv bewertet haben. Insgesamt war auffallend, dass sich alle Probanden am Ende der Befragung über den geplanten weiteren Verlauf des Trainingsassistenten erkundigten. Bei allen Befragten bestand also auch nach der Testphase ein Interesse an TASI.

Bezüglich des vorliegenden Forschungskontexts lässt sich aus der Beobachtung schlussfolgern, dass eine positive Nützlichkeit (Mehrwert) und eine einfache Bedienbarkeit wahrgenommen wurde. Auch wurde bei den Probanden eine positive Einstellungs- und Verhaltensakzeptanz gegenüber TASI beobachtet. Es konnten somit

über alle drei Felder (Nützlichkeit, Bedienbarkeit, Akzeptanz) Erkenntnisse gewonnen werden. Die Zusammenhänge müssen in einer abschließenden Betrachtung erörtert werden.

8.3 Resümee

Im Bezug zum Forschungskontext, mit dem Ziel der Konzeption, Entwicklung und Evaluation eines mobilen Trainingsassistenten auf Basis von Web-Technologien für den Bereich Handball, konnten die erstellten Hypothesen bestätigt und die Forschungsfragen beantwortet werden. Damit kann man das Forschungsprojekt aber noch nicht als beendet betrachten. Aus Sicht der DBR-Methode ist der Design-Prozess ein ständiger Kreislauf. Der vorliegende Prototyp kann weiterentwickelt werden und es können neue Forschungsfragen und Hypothesen gebildet werden.

9 Abschließende Diskussion

In einem Prozess der Vorprüfung, Konzeption und Umsetzung wurde der Prototyp TASI unter ständigem Re-Design entwickelt und getestet. Hinsichtlich der Problemstellung kann aufgrund der ausgewerteten Ergebnisse bestätigt werden, dass mobile Web-Technologien, wie der Trainingsassistent TASI, im Sport gewisse Potentiale besitzen und zur Bewältigung bestimmter Problemfelder beitragen können. Der im Rahmen der Forschungsarbeit betrachtete wissenschaftliche Kontext bezüglich des Trainingsassistenten generiert weiteren Forschungsbedarf. Wird der Blick auf den DBR-Ansatz gerichtet, ist zum einen eine Weiterentwicklung des Prototyps notwendig. Zum anderen können auf Grundlage der gewonnenen Forschungsergebnisse weitere Forschungsfragen abgeleitet werden.

Die nachfolgend genannten Themenbereiche sollten bei einer Weiterentwicklung von TASI beachtet werden: Die beobachteten Usability-Probleme können durch Hinzunahme bestimmter Technologien, wie z.B. Ajax (ein Konzept zur asynchronen Datenübertragung zwischen Browser und Server) minimiert werden. Der Funktionsumfang von TASI kann hinsichtlich der durchgeführten Anforderungsanalyse erweitert werden. Diesbezüglich sind tiefgehende Usability-Test, im Rahmen von Anwendungstests, mit Protokollierung des Nutzungsverhaltens in Nutzungstests um weitgehende Informationen zu erhalten, zu empfehlen. In diesem Zusammenhang ist auch eine didaktische Bewertung der Übungsvideos möglich. Bei der Erstellung von Übungsvideos ist ein Demonstrator, der den erforderlichen Bewegungsablauf bestmöglich vollbringen kann, das wichtigste Kriterium. Die Beschreibung der Videos müssten hinsichtlich der Verbesserungsvorschläge der Befragten optimiert werden (die ungefähre Durchführungsdauer der Übung soll dem Nutzer helfen das Training besser planen zu können. Eine Altersfreigabe für die Übungen sollte beigefügt werden. Zudem sollte in der Übungsbeschreibung ein Vorschlag zur Wiederholungs- und Serienanzahl gegeben sein). Sicherheitsstandards bezüglich der Web-Entwicklung wurden im

bisherigen Prototyp nicht weiter beachtet. Diese sind für größere Untersuchungen ggf. essentiell. Neben dem Bereich der Athletik könnten auch Übungen aus den Bereichen Technik und Taktik in den Trainingsassistenten aufgenommen werden. Vor allem, der von den Befragten erwähnte Vorschlag, Torwartraining in TASI zu implementieren, ist sinnvoll. In diesem Bereich weisen die meisten Trainer gewisse Defizite auf. Die Torwartposition ist im Handball, gegenüber dem Feldspieler, gesondert zu betrachten. In vielen Trainingseinheiten wird genau das vernachlässigt. Der Trainingsassistent kann auch sportartübergreifend untersucht werden. Vor allem aufgrund der Tatsache, dass die Forschungsergebnisse das bevorzugte Einsatzgebiet von TASI auf individuelles Training oder Kleingruppenttraining gelegt haben, kann er so in fast jeder Sportart eingesetzt werden (entsprechende Anpassungen vorausgesetzt).

Im Bereich Handball hat sich der Trainingsassistent auf Grund der Problemstellung und der durchgeführten Tests im Bereich des individuellen Trainings empfohlen. Jedoch muss an dieser Stelle die Art des Einsatzes kritisch betrachtet werden. Die aktuellen Übungsvideos sind zum Großteil im Athletikbereich angesiedelt. Einige Übungen zielen auch auf ein Techniktraining ab. Der Trainingsassistent kann dem Nutzer zwar den korrekten Bewegungsablauf unter Berücksichtigung der einblendeten Grafiken und Erklärungen detailliert schildern, jedoch kann er den Nutzer in der Ausführung nicht korrigieren. Somit kommt es immer auf die Übung an, ob der Trainingsassistent vom Nutzer alleine oder unter Anwesenheit des Trainers benutzt werden kann. Diesbezüglich könnten auch weitere Forschungsthemen erörtert werden. Eventuell ist es möglich, ein Videofeedback-Training in den Assistenten zu implementieren.

Der Bereich der Akzeptanzforschung bringt ebenfalls Forschungspotential mit sich. Die in der Arbeit untersuchte Verhaltens- und Einstellungsakzeptanz, im Sinne der tatsächlichen Nutzung ist von mehreren Variablen als den Getesteten (PU und PEoU), abhängig. Dazu sollten weitere Untersuchungen unter Anbetracht anderer Modelle, wie beispielsweise dem TAM2 von Venkatesh & Davis [VD00], unternommen werden. Diesbezüglich sind Befragungen in größerem Umfang sicherlich sinnvoll.

Zusammenfassend zeigt sich, dass weiterführende Arbeiten, vor allem in den Bereichen Medienpädagogik, Mediendidaktik, Informatik und Sport, zu vollziehen sind. Untersuchungen können in verschiedene Richtungen erfolgen. Vor allem die Weiterentwicklung des Prototyps erscheint aufgrund des relevanten Forschungskontexts als sinnvoll.

Literaturverzeichnis

- [BM04] Bürg, O.; Mandl, H.: Akzeptanz von E-Learning in Unternehmen. Forschungsbericht 167. München: Institut für Pädagogische Psychologie, Ludwig-Maximilians-Universität, 2004.
- [Br96] Brooke, J: SUS - A 'quick and dirty' usability scale. In (Jordan, P. W.; Thomas, B.; Weerdmeester, B. A.; McClelland, I. L., Hrsg): Usability Evaluation in Industry, Taylor & Francis, London, UK, S. 189-194, 1996.

- [Co04] Cooper, A.: *The Inmates Are Running the Asylum: Why High-tech Products Drive Us Crazy and How to Restore the Sanity*. Sams Publishing, Indianapolis, IN, USA, 2004.
- [CR90] Carroll, J. M.; Rosson, M. B.: Human-computer interaction scenarios as a design representation. In: *Proceedings of the 23rd Annual Hawaii International Conference on Systems Sciences*, HI, Kailua-Kona, S. 555-561, 1990.
- [DBW89] Davis, F. D.; Bagozzi R. P.; Warshaw, P. R.: User acceptance of computer technology: A comparison of two theoretical models, *Management Science* 35, S. 982-1003, 1989.
- [De03] Design-Based Research Collective: *Design-Based-Research: An Emerging Paradigm for Educational Inquiry*. *Educational Researcher*, 32/1, S. 5-8, 2003.
- [FTB05] Fischer, U.; Thienes, G.; Bredel, F.J.: CD-ROMs für den Sportunterricht und die Sportlehrererausbildung – ausgewählte Evaluationsergebnisse. *Sportunterricht* 54/1, S. 11-16, 2005.
- [HHM09] Herrington, A.; Herrington, J.; Mantei, J.: Design principles for mobile learning. In (Herrington, J.; Herrington, A.; Mantei, J.; Olney, I.; Ferry, B., Hrsg.): *New technologies, new pedagogies: Mobile learning in higher education*. University of Wollongong, Australia, Wollongong, S. 129-138, 2009.
- [HKV13] Hebbel-Seeger, A.; Kretschmann, R.; Vohle, F.: *Bildungstechnologien im Sport. Forschungsstand, Einsatzgebiete und Praxisbeispiele*. In (Ebner, M.; S. Schön, S., Hrsg.): *L3T: Lehrbuch für Lernen und Lehren mit Technologien* (2. Auflage), 2013.
- [Ho96] Hohmann, A.: *Trainingssteuerung im Sportspiel*. Ahrensburg: Czwalina, 1996.
- [Ih91] Ihlo, H.: DDR – Bildungsmedien im Sportunterricht. In (Dannenmann, F., Hrsg.): *Volleyball gesamtdeutsch*. 16. Symposium des Deutschen Volleyballverbandes 1990. Ahrensburg, S. 222-229, 1991.
- [Kr15] Kromer, A.: *Positionstraining für Rückraum-, Kreis- und Außenspieler*. Philippka Sportverlag, Münster, 2015.
- [Le16] *Learntec: Digitale Schule - vernetztes Lernen 2016*, <https://www.bitkom.org/Presse/Anhaenge-an-Pls/2016/Charts-Digitale-Schule-13-01-2016-final.pdf>, Stand: 14.05.2016.
- [Re05] Reinmann, G.: Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtswissenschaft*, 33/1, S. 52-69, 2005.
- [Re14] Reinmann, G.: *Design-based Research: Auftakt für eine methodologische Diskussion entwicklungsorientierter Bildungsforschung 2014*, http://gabi-reinmann.de/wp-content/uploads/2013/05/Reader_Entwicklungsforschung_Jan2015.pdf, Stand: 20.07.2016.
- [VD00] Venkatesh, V.; Davis, F. D.: A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* 46/2, S.186-204, 2000.

Hybride Lernräume für Innovationsprozesse

Christian Kohls¹, Guido Münster²

Abstract: Dieser Beitrag beschäftigt sich mit der Planung und Gestaltung hybrider Lernräume für Innovationsprozesse mithilfe einer Entwurfsmustersprache. Zunächst werden strategische und operative Ziele für die Gestaltung von Innovationsräumen an Hochschulen vorgestellt. Basierend auf existierenden Frameworks für die Gestaltung von Innovationsräumen sowie einer induktiv-empirischen Analyse von existierenden Good Practices und Technologien werden Entwurfsmuster entwickelt und in Kurzform vorgestellt. Die Entwurfsmuster haben bei der konkreten Planung dreier Innovationsräume am Campus Gummersbach der TH Köln einen wichtigen Beitrag geleistet, um Design-Entscheidungen zu treffen und ungewöhnliche Lösungsansätze zu vermitteln. Sie können für ähnliche Projekte an anderen Hochschulen wichtige Gestaltungshinweise geben.

Keywords: Design Patterns, Soziotechnische Systeme, Lernräume, Innovationswerkzeuge, Design Thinking

1 Einleitung

Unter hybriden Lernräumen (auch Blended Spaces) wollen wir im Folgenden Bildungsräume verstehen, in denen sich verschiedene Räume überblenden, z.B. der physikalische, digitale, informationelle, konzeptionelle und der soziale Raum [Be14]. Die Überblendung der Räume zielt nicht nur auf das Aufsummieren bestehender Funktionalitäten einzelner Räume ab (nicht nur „best of both worlds“), sondern eröffnet neue Anwendungsszenarien für Lern- und Lehrsituationen („neue Welten“). Für den Informatikcampus an der TH Köln haben wir Räume konzipiert, mit denen sich innovative Denkprozesse unterstützen lassen. Insbesondere werden die verschiedenen Phasen des Design Thinkings begünstigt. Wir wollen Raum für Innovation schaffen und dabei die verschiedenen Aktivitäten während eines Designprojekts sowohl mit analogen als auch mit digitalen Werkzeugen fördern. Wichtig scheint uns dabei eine nahtlose Verschmelzung beider Welten, z.B. indem digitale Artefakte auf greifbare physikalische Objekte gemappt werden und andersherum analoge Arbeitsergebnisse direkt und ohne Brüche digitalisiert und innerhalb digitaler Umgebungen weiter manipuliert werden können. Für die Planung solcher Arbeits- und Lernumgebungen gibt es bereits verschiedene Ansätze und Frameworks, die bei der Entscheidungsfindung für bestimmte Technologien und Gestaltungsoptionen helfen. Aus unserer Sicht sind diese Frameworks jedoch häufig auf einer zu abstrakten Ebene angesiedelt, so dass sie zwar wichtige Kriterien definieren, aber nicht immer handlungsorientierte (und empirisch überprüfbare) Empfehlungen bereitstellen. Als Alternative sehen wir daher die

¹ TH Köln, Institut für Informatik, Steinmülleralle 1, 51643 Gummersbach, christian.kohls@th-koeln.de

² TH Köln, Institut für Informatik, Steinmülleralle 1, 51643 Gummersbach, guido.muenster@th-koeln.de

Entwicklung einer Entwurfsmustersprache, mit der gute Praktiken bei der Gestaltung hybrider Lernräume erfasst werden. Der Entwurfsmusteransatz ist in der Informatik insbesondere im Bereich der Softwareentwicklung etabliert [Ga95], findet jedoch auch in vielen anderen Bereichen Einsatz [Ko14], u.a. im Interaktionsdesign [Bo01] sowie zur Gestaltung soziotechnischer Systeme [SL07]. Der Ansatz kommt ursprünglich aus der Architekturtheorie [Al77] und scheint daher gerade für die Gestaltung von Räumen besonders geeignet. Dieser Artikel stellt ausgewählte Entwurfsmuster vor und schildert unsere Erfahrungen bei der Umsetzung der Räumlichkeiten mithilfe einer Mustersprache. Zunächst sollen jedoch die Ziele der Räume skizziert und existierende Frameworks vorgestellt werden. Daraus leiten sich dann Entwurfsmuster ab. Das Fazit geht auf unsere Erfahrungen bei der Verwendung der Mustersprache für die konkrete Raumplanung und Umsetzung ein.

2 Ziele bei der Gestaltung von Innovationsräumen

Mit den hybriden Innovationsräumen sind bestimmte strategische und operative Ziele verknüpft, die sich sowohl auf den Hochschulkontext als auch auf die gesellschaftliche Entwicklung beziehen – z.B. das Ermöglichen von sozialer Innovation.

2.1 Strategische Ziele

Auf strategischer Ebene hat der Raum vor allem das Ziel, die Kreativität und das Innovationspotential studentischer Projekte zu erhöhen. Ein hybrider Innovationsraum soll Methodenkompetenz vermitteln, Ort der interdisziplinären Begegnung sein und Design Thinking unterstützen. Er soll zudem die Identifikation mit der eigenen Hochschule intensivieren, indem diese als Ort der Innovation erlebbar gemacht wird. Heutzutage sind kooperative Arbeitsformen der eigentliche Anlass für die Begegnung auf dem Campus, da Lehr- und Lernmaterialien zur Wissensvermittlung digital von zuhause aus und unterwegs nutzbar sind. Der Raum soll daher etwas Besonderes sein.

Als weiteres strategisches Ziel aus Informatiksicht kann die Entwicklung neuer Geräte und Anwendungen für kreative und kollaborative Arbeitsformen genannt werden. Der Raum soll zum Beispiel im Rahmen von Bachelor-, Master- und Doktorarbeiten eine „Spielwiese“ zum Ausprobieren neuer Apps, Widgets und Gadgets bieten. Als drittes strategisches Ziel sehen wir die empirische Erprobung von Raumkonstellationen und digitalen Werkzeugen. Die Fragestellungen hierzu lauten z.B.: Welchen Einfluss haben unterschiedliche interaktive Technologien gegenüber klassischen Moderationsmaterialien? Wie muss eine Kreativitäts-App gestaltet sein, damit sie in möglichst vielen Szenarien effektiv genutzt werden kann?

2.2 Operative Ziele

Auf der operativen Ebene geht es darum, wie der Raum Lehr- und Lernszenarien beeinflusst und wie der Raum aufgesetzt werden muss, damit er einerseits als experimentelle „Spielwiese“ und gleichzeitig als alltagstauglicher Arbeitsraum genutzt werden kann. Zudem gilt es, eine entwicklungsbasierte und empirische Forschung zu ermöglichen. Auf der operationalen Ebene fördert der Raum projektbasiertes und kompetenzorientiertes Lernen. Er etabliert eine ergebnisorientierte „Maker-Kultur“ und unterstützt spielerische Ansätze bei der Entwicklung oder Vermittlung ernsthafter Inhalte.

2.3 Bedeutung für Hochschulentwicklung

Über die Bedeutung der Raumgestaltung wird im hochschuldidaktischen Kontext reflektiert [DI13]. Skerlak, Kaufmann und Bachmann [SKB14] sagen über Lernumgebungen an der Hochschule, dass zunehmend Lernräume für selbstgesteuertes Lernen entstehen, Zwischenräume für informelles Lernen genutzt werden und Spielräume für das Entwickeln von innovativen Lehr- und Lernformen entstehen. Der Horizon-Report [Ad17] hat in den letzten Jahren immer wieder die Themen „Makerspace“, „Bring your own device“, „Flipped Classroom“, „Internet of Things“, „Wearables“ und „Redesigning learning spaces“ aufgegriffen. Eine Auseinandersetzung mit diesen Konzepten sollte von hochschulstrategischer Bedeutung sein. Auch die Expertenkommission des Stifterverbands kommt in „The Digital Turn“ [Ho16] zu dem Schluss, dass „Makerspaces und kreative Räume“ sowie „digitale Kollaborationstools“ zu den Schlüsseltechnologien der digitalen Hochschule zählen.

2.4 Design Thinking ermöglichen

Ein zentrales Ziel der Innovationsräume ist es, die Prozesse des Design Thinkings, insbesondere die Phasen der Ideation (also der Ideenfindung) und des Rapid Prototypings zu unterstützen. Design Thinking ist ein an der Stanford School of Engineering entwickeltes Konzept für Design Prozesse [LNM09]. Ziel des Design Thinkings ist das systematische Ermöglichen von Innovation [MT16]. Es handelt sich um einen iterativen Prozess, der nutzer- und kundenzentrierte „Ergebnisse zur Lösung komplexer Probleme[n] liefert“ [Ue15]. Der Design-Thinking-Prozess besteht sowohl aus analytischen Phasen (Sammeln, Ordnen, Bewerten) als auch aus synthetischen Phasen (Entwickeln, Erproben, Verbessern) [PMW09]. Ein zentrales Element des Design Thinkings ist das frühe Prototyping. Es wird davon ausgegangen, dass vor allem am konkreten Modell – auch wenn dieses nur rudimentär umgesetzt ist – besonders gut Schwachstellen und Weiterentwicklungspotentiale erkannt werden. Skizzen, Zeichnungen und einfachste Modelle erlauben das gleichzeitige Explorieren von Problem- und Lösungsraum [Cr11]. Wichtig bei der Raumplanung ist daher, dass vielfältige Werkzeuge und Materialien für das Entwickeln und Ausprobieren neuer Dinge bereitstehen. In unserem Fall sollen

digitale und analoge Elemente miteinander verknüpft werden, um reichhaltige und neuartige Arbeitsformen zu ermöglichen.

3 Hybride Räume

Ein hybrider Raum bzw. ein Blended Space entsteht, wenn z.B. digitale und physische Bereiche eng miteinander verbunden werden und sich überlagern [Be14]. Das Ergebnis dieser Vermischung ist ein neuer Bereich, der ganz eigene, neue emergente Eigenschaften besitzt und somit mehr ist als die Summe der Einzelteile.

3.1 Blended Spaces

Der Blended Space Ansatz findet in verschiedenen Bereichen Anwendung: interaktive Touristenführer [Be14], „Living History“ Bereiche in Museen oder Interactive Collaborative Environments [BM16]. Seinen Ursprung findet der Ansatz in der Blending Theory von Fauconnier und Turner [FT08], die von Benyon [Be14] auf Blended Spaces angewandt und weiterentwickelt wird. Benyon und Mival zeigen verschiedene Räume oder Bereiche auf, die für die Blended Theory von Bedeutung sind [BM15]. Dazu gehören der physikalische Raum, der digitale Raum, der Informationsraum, der Konzept- und Begriffsraum, der Navigationsraum und schließlich der Sozialraum. Die Aufgabe eines Designers bei der Vermischung eben genannter Räume zu einem Blended Space besteht darin, den neu zu gestaltenden Bereich so natürlich und intuitiv wie möglich zu erschaffen.

3.2 Frameworks zur Raumgestaltung

Der Blended Space Ansatz bietet bereits einen guten theoretischen Rahmen, um Designprinzipien abzuleiten. Darüber hinaus gibt es verschiedene Frameworks, die sich mit der Gestaltung von Innovationsräumen oder mit Blended Spaces beschäftigen. So geht das TACIT-Framework direkt aus dem Blended Space Ansatz hervor [BM15], indem Schlüsselemente für die Raumgestaltung abgeleitet werden. Dazu gehören Territorium des Raums, die Bewusstwerdung der Möglichkeiten, die Steuerbarkeit und Interaktionsformen sowie die verschiedenen Übergänge (z.B. Austauschen von Dateien, Sammeln von Arbeitsergebnissen), die innerhalb des Blended Raums möglich sind.

Moultrie et al. [Mo07] entwickeln ein Framework, das weniger Aussagen über die Gestaltung eines Innovationsraums trifft, als vielmehr über die damit verbundenen Ziele. Als Ziele nennen sie strategische und symbolische Ziele, Innovationseffizienz und Effektivität, verbesserte Teamarbeit, stärkeres Einbeziehen von Stakeholdern sowie allgemein den Ausbau unternehmerischer Fähigkeiten. Das Framework operationalisiert diese Ziele und differenziert zwischen Faktoren zur Planung des Raumes sowie Faktoren zur Bewertung des Zielerreichungsgrads.

Bustamante et al. [Bu15] vergleichen die Kriterien verschiedener Frameworks zur Gestaltung von Innovationsräumen. Allen Frameworks ist gemein, dass sie wichtige Prinzipien, Faktoren, Kriterien und Richtlinien für die Planung und ggf. spätere Überprüfung nennen. Wenn es jedoch um die konkrete Gestaltung eines neuen Raums geht, so bleiben sie sehr unspezifisch und häufig zu abstrakt. Eine Alternative zu allgemeinen Frameworks stellt daher aus unserer Sicht der Entwurfsmusteransatz nach Christopher Alexander [Al77], [Al79] dar.

4 Entwurfsmuster für die Gestaltung

Entwurfsmuster erfassen erprobte Lösungen in generischer Form und vermitteln Designwissen. Sie sind eine Ebene spezifischer als Frameworks, da sie für bestimmte Situationen und Kontexte konkret benennen, welche Lösungsformen angemessen sind und welche Konsequenzen bei der Umsetzung zu berücksichtigen sind. Sie sind jedoch gleichzeitig allgemein genug, um eine Anpassung auf die spezifischen Kontexte einer Raumgestaltung zu berücksichtigen. Der Entwurfsmusteransatz erscheint daher vielversprechend, um die Designoptionen auf eine konkretere und besser fundierte Ebene zu heben.

4.1 Empirische Grundlage: Pattern Mining

Das Identifizieren und Dokumentieren von Entwurfsmustern wird Pattern Mining genannt. Es gibt verschiedene Vorgehensweisen für das Pattern Mining [KP09]. Der hier gewählte Ansatz ist ein typischer Methodenmix aus induktiver und deduktiver Vorgehensweise:

- Ableiten von Anforderungen und Einflussfaktoren aus den Frameworks
- Analyse von existierenden Good Practice Beispielen zur Raumgestaltung
- Besuch von Messen, Sichtung von Prospekten, Besichtigung von Räumen
- Testen von analogen und digitalen Werkzeugen
- Partizipative Designsessions mit späteren Nutzern des Raums (Studierende)
- Evaluation verschiedener Gestaltungsmöglichkeiten durch Mockups

Bei der Formfindung für die Muster können die Erkenntnisse aus den Raumgestaltungsframeworks als Einflussfaktoren berücksichtigt werden. Zudem kann für das Auffinden wiederkehrender Strukturen auf Fallstudien über existierende Räume zurückgegriffen werden: i-Land [St99], ICE [BM12], Makerspaces [Fo14] und iRoom [JFW02]. Weiterhin wurden zahlreiche analoge Werkzeuge (z.B. Methodenkarten, Innovation Games) und digitale Werkzeuge (z.B. iPads mit Stift, Touchscreens, Tabletops) im

Labor oder bei Messebesuchen getestet. Verschiedene Gestaltungsoptionen für den Raum wurden durch Mockups durchgespielt, z.B. um die Größenverhältnisse für einen zentralen Arbeitstisch einschätzen zu können.

Aus der Analyse haben sich verschiedene Musterkandidaten ergeben. Musterkandidaten, die durch mehrere Beispiele belegt und mithilfe eines Theorierahmens plausibel erklärt werden können, sind in die engere Auswahl gekommen. Durch ein Clustering nach Problemkategorien sind verschiedene Gruppen entstanden: Arbeitsmaterial und Werkzeuge, Atmosphäre, Sessionmanagement, Navigation und Blended Interaction.

4.2 Vielfältige analoge und digitale Materialien

Die erste Kategorie bilden die verfügbaren Arbeitsmaterialien. Da nicht immer genau absehbar ist, welche Materialien gerade benötigt werden und es auch sehr wahrscheinlich ist, dass ungewöhnliche Materialien neue Wege eröffnen können, sollte immer eine große Menge an unterschiedlichen Materialien zur Verfügung stehen. Beispiele sind Boxen mit Arbeitsmaterialien für Design Thinking [Ue15] oder Schränke mit Arbeitsmaterialien bei der Design Agentur IDEO [KL01]. Dies führt zu dem Muster ABUNDANCE OF MATERIALS, welches weiter verfeinert wird. BUILDING MATERIALS bezieht sich auf Baumaterialien, z.B. Legosteine für Serious Play [KR14] oder Materialien zum Tüfteln (Tinkering). THOUGHT TRIGGERS sind im Raum verfügbare Impulsgeber für Ideen, z.B. Zufallsbilder, Story-Würfel, Reizwörter – sowohl in analoger als auch in digitaler Form. Für die Visualisierung stehen A BUNCH OF PEN AND PAPERS ergänzend zu digitalen Zeichentablets und interaktiven Whiteboards bereit. Unter CURIOUS THINGS verstehen wir ungewöhnliche Objekte, die zum Querdenken anregen sollen. Diese können offen herumliegen oder auch als Überraschungselement in Behältern versteckt sein (SURPRISE). Weiterhin sollten für analytische Prozesse verschiedene Arbeitsvorlagen (WORK TEMPLATES, z.B. Business Model Canvas) bereitgestellt werden. Die ausgefüllten Vorlagen sollten direkt digitalisierbar sein und eventuell weitere Aktionen automatisch anstoßen (TEMPLATE TO ACTION), z.B. automatisches Einsortieren in verschiedene digitale Listen (etwa Pro und Kontra), Verwenden von Zeichnungen innerhalb anderer Programme (Spiele, Living Mockups).

4.3 Atmosphäre

Die Atmosphäre des Raumes spielt ebenfalls eine besondere Rolle bei der kreativen Arbeit. Daher haben wir auch in diesem Bereich zahlreiche Muster identifiziert, die sich vor allem im physikalischen Raum bewegen, z.B. das allgemeine Ambiente, Pflanzen, die Ausstattung mit hochwertigen Möbeln, Stehtische und Werkbänke, Sofas, Bilder und Sound. Die Ausstattung mit digitalen Werkzeugen bringt jedoch neue Probleme mit sich und erfordert weitere Lösungsansätze. Dazu gehören die Muster NO CABLE (möglichst keine sichtbaren Kabel), Steuerung von Sound und Illumination (z.B. Beleuchten von bestimmten Arbeitsbereichen, atmosphärische Beleuchtung). In der Kategorie „Food &

Convenience“ sind zudem Muster gefunden worden, die dem allgemeinen Wohlbefinden dienen, z.B. die Bereitstellung von Snacks und insbesondere die Erlaubnis, im Raum bedenkenlos Getränke und Snacks zu verzehren. Dies ist insofern wichtig, da die technologische Ausstattung nicht dazu führen sollte, dass Nutzer*innen keinen Kaffeebecher mehr auf den (interaktiven) Tisch abstellen mögen (FOOD IS NO HAZARD). Weiterhin bedarf es eines Raumkümmerers (ROOM CONCIERGE), der dafür Sorge trägt, dass für die nächste Sitzung alle Werkzeuge bereitstehen.

4.4 Session Management

Unter Session Management haben wir Muster zusammengefasst, die es ermöglichen, innerhalb des digitalen Raums Sitzungen zu sichern, in andere (physikalische) Räume mitzunehmen und die Arbeit an einer bestimmten Stelle wieder aufnehmen zu können.

Zur Datensicherheit muss eine Backupfunktionalität vorhanden sein. Um verlorene Daten wieder zurückzuholen, muss eine Wiederherstellungsfunktionalität gegeben sein. Das Rollback versetzt die Anwendungen und Systeme wieder auf einen Basiszustand zurück (BACKUP – RECOVERY – ROLLBACK). Dies sind oft Standardfunktionen bei Onlinewerkzeugen wie Google Docs oder Wikipedia. Dennoch halten wir die explizite Dokumentation als Muster für wichtig, nämlich für die Konzeption neuer Werkzeuge.

Mit ROOM-RESET ist gemeint, dass sich alle Systeme wieder in ihren Ausgangszustand zurücksetzen, z.B. damit die nächste Arbeitsgruppe im Raum nicht auf die Ergebnisse anderer Gruppen zugreifen kann. Ein Room-Reset-Feature findet sich z.B. innerhalb der Software von interaktiven Whiteboards (etwa SMART Boards, Microsoft Surface Hub).

Wenn eine Gruppe von Nutzern den Raum eine Woche oder einen Monat später erneut belegt, um an der gleichen Stelle weiterzuarbeiten, an der sie aufgehört hatte, muss es möglich sein, über eine Historie auf diese zuzugreifen. Alle Geräte im Raum werden auf diesen Stand zurückversetzt und bieten den Gesamtzustand der letzten Veranstaltung zur Bearbeitung an (ACCESS TO OLD SESSIONS). Da es mehrere unabhängige Innovationsräume gibt, müssen die Sessions auch mit in andere Räume genommen werden können. Dies ist auch für Break-out-sessions relevant. Dies führt zum Muster CROSS-ROOM-INTERACTION.

4.5 Navigation

Auch der Navigationsraum muss in einem hybriden Raumkonzept berücksichtigt werden. Hier gibt es verschiedene Lösungen, die Nutzer*innen dabei helfen, sich im Raum zurecht zu finden – z.B. können so bestimmte analoge und digitale Werkzeuge ad-hoc erklärt werden. In einem Raum, der viele Möglichkeiten bietet, kann nicht alles auf den ersten Blick ersichtlich drapiert werden oder intuitiv benutzbar sein. Zur Orientierung und Anleitung dienen TANGIBLE INSTRUCTIONS, z.B. als Menü-Karten aufgebaute Anleitungen, kurze Übersichtspläne oder Kochbücher mit Zutatenlisten und

Vorgehensweise zum Zusammenführen der einzelnen Bausteine. Über QR-Codes können diese physikalischen Elemente wiederum Videoanleitungen Smartphones oder Wanddisplays aufrufen.

Die Oberflächen der digitalen Geräte müssen so gestaltet sein, dass auf den ersten Blick ersichtlich ist, was ein Icon auslöst, so dass der Benutzer dazu ermuntert wird, dieses auch zu drücken. Objekte wie Legosteine, die im Raum vorgehalten werden, müssen so drapiert sein, dass sie dazu einladen, sie in die Hand zu nehmen und etwas damit zu generieren. Das entsprechende Muster ist INVITE ACTION.

Die Muster DEFAULT APPS und FAST ACCESS beschäftigen sich damit, dass ein möglichst schneller Zugang auf die wichtigsten Funktionen und Werkzeuge gegeben sein muss. Standardanwendungen sollten so zugreifbar sein, dass nicht lange nach ihnen gesucht werden muss. Tablets haben z.B. keine Bootzeiten, Systeme sollten innerhalb der Kernzeiten vorab eingeschaltet sein oder sehr schnell hochfahren.

4.6 Blended Interaction

Vielleicht am interessantesten ist die Kategorie der Blended Interaction, da sie sich mit Lösungen befasst, die eine nahtlose Verknüpfung der analogen und digitalen Welt ermöglichen sollen.

Das Muster PHYSICAL TO DIGITAL bewirkt, dass physische Handlungen und getätigte Aktionen sofort digitalisiert werden können. Wenn ein Benutzer mit einem Stift auf eine digitale Oberfläche oder einem Blatt Papier etwas skizziert, dann soll es sofort durch eine berührungsempfindliche Oberfläche oder eine Kamera digitalisiert werden. Beispiele sind die Aufnahme von Bildern mit Smartphones oder Dokumentenkamera. Handgeschriebene Einträge mittels Stift auf einem berührungsempfindlichen Medium werden umgehend digitalisiert und als 1:1 Abbildung oder umgewandelt in Text zur weiteren Bearbeitung zur Verfügung gestellt.

Das komplementäre Gegenstück ist das Muster DIGITAL TO REAL. Im digitalen Raum erstellte Artefakte sollen nicht nur digital bleiben, sondern in die reale Welt geholt werden. Beispiele sind Projektionen auf physikalische Objekte, an Wände oder auf den Boden, das Drucken mit Foto- oder 3D-Drucker, sowie Steuerung von Robotern oder Raspberry Pi Modulen.

Physische Objekte werden zudem genutzt, um mit ihrer Hilfe eine Interaktion zwischen digitaler und analoger Welt auszulösen (PHYSICAL OBJECTS TRIGGER ACTIONS). Ein Beispiel ist das Digitalisieren einer Skizze durch Druck auf einen physischen Knopf.

Enthält ein Raum mehrere elektronische Geräte, so sollten sie miteinander verbunden werden, um als eine Einheit zu fungieren (DEVICE ORCHESTRATION). Beispiele sind das Senden von Brainstorming-Beiträgen via Smartphone ans interaktive Whiteboard, Skizzieren von Labyrinthen am Whiteboard für Spiele auf einem Tablet oder einer

Spiele-Arcade, die Auswahl von Inhalten für große Displays über kleine interaktive Wandbilder.

Einen Schritt weiter geht der Zusammenschluss mehrerer Geräte zu einer größeren Einheit (COUPLED DEVICES). Dies ermöglicht z.B. eine größere zusammenhängende Arbeitsfläche. Die Geräte können alleine oder als ein Ganzes genutzt werden. Beispiele sind Doppelprojektion interaktiver Whiteboards, die über einen Rechner gesteuert werden, kleine Infoscreens, die gemeinsam eine große Fläche ergeben, eine Wandtrilogie mit interaktiven Tablets, die zusammenspielen, oder mehre Tablets auf dem Tisch, die ein großes Tabletop temporär abbilden.

Das Einbeziehen selbstmitgebrachter Geräte (BRING YOUR OWN DEVICE) ist ein weiteres Entwurfsmuster. Studierende oder Mitarbeiter nutzen ihre privaten Geräte zur Eingabe oder für Abstimmungen (z.B. Audience Response Systeme). Damit die selbstmitgebrachten Geräte nicht ständig am Stromnetz hängen müssen (NO CABLE), aber dennoch aufgeladen werden können, sind LADESTATIONEN bereitzustellen. Beispiele sind versteckte Ladekabel in Boxen, Ladesäulen auf Flughäfen oder in Hotels.

Für die Steuerung der Geräte sollten auch die Eingabemöglichkeiten weitergedacht werden. Neben den handschriftlichen Eingaben und der Steuerung durch tangible Objekte ist eine Sprachsteuerung eine Option. Es eröffnet völlig neue Wege der Interaktion, wenn die Systeme zuhören und Gesagtes aufgreifen, um eine bestimmte Aktion auszuführen. Statt nach einer Vorlage für eine Stärken-Schwächen-Analyse auf dem Computer zu suchen, kann dem Raum einfach gesagt werden, dass diese geöffnet werden soll: „Raum, öffne SWOT-Vorlage“. Auch die Ausrichtung und Nutzung von Möbelstücken kann als Input für die digitale Welt dienen (FURNITURE AS INPUT). So kann gezählt werden, wie häufig Gegenstände genutzt werden oder welche Stühle beieinanderstehen. Die Ausstattung von Möbelstücken mit Bewegungs-, Gewichts- und Ausrichtungssensoren lässt eine Analyse der Raumaktivitäten zu, kann aber auch für spielerische Aktivitäten genutzt werden.

5 Fazit

Die identifizierten Entwurfsmuster konnten zum großen Teil für die von uns konzipierten Innovationsräume umgesetzt werden und haben sich als sehr gutes Planungswerkzeug erwiesen. Die Entwurfsmuster beruhen auf der Beobachtung, Analyse und dem Experimentieren mit existierenden Lösungen. Sie ersetzen dabei die eingangs erwähnten Frameworks nicht, aber machen diese eine Stufe konkreter und damit empirisch überprüfbarer. Die Frameworks liefern zudem einen theoretischen Bezugsrahmen für die Begründung der Lösungen.

Bei der Planung haben uns die Muster einerseits geholfen, konkrete Technologien hinsichtlich der von uns geforderten Eigenschaften zu überprüfen. Die durch Muster geschaffene Abstraktionsebene hilft dabei, konkrete Implementierungsentscheidungen

zu treffen und sie zu begründen. Durch die Erörterung der Konsequenzen eines Entwurfsmusters werden nicht nur die positiven Effekte, sondern auch mögliche Einschränkungen transparent. So gibt es einige Muster, die negative Auswirkungen auf das Sicherheitsempfinden der Nutzer*innen haben können, z.B. die automatische Sprachanalyse sowie die zahlreichen Kameras im Raum. Durch die explizite Nennung von Folgeproblemen kann nach weiteren Lösungsansätzen gesucht werden.



Abb. 1: ideenreich – Innovationsraum mit analogen und digitalen Werkzeugen

Muster helfen auch dabei, die Elemente des Raums besser an andere Stakeholder zu vermitteln. Durch den guten Begründungsrahmen wird es für Raumplaner z.B. einfacher, ungewöhnliche Elemente des Raums (etwa der Einsatz von Legosteinen oder Innovationskartenspielen) besser zu vertreten.

Weiterhin haben die Muster dabei geholfen, eine komplexe Raumkomposition in einzelne Bestandteile aufzubrechen, die unabhängig betrachtet werden können und dennoch in das Gesamtkonzept passen. Das harmonische Zusammenspiel ist dabei über Querbezüge zwischen den Mustern gewährleistet.

Teils decken die Entwurfsmuster auch technologische Defizite auf. So ist das Entwurfsmuster DEVICE ORCHESTRATION zwar keine Utopie, sondern aus konkreten Beispielen der Gerätekopplung empirisch abgeleitet. Das Zusammenspiel aller digitalen Geräte im Raum ist derzeit jedoch noch nicht vollständig umgesetzt. DEVICE ORCHESTRATION bleibt aus unserer Sicht dennoch eine valide Lösung, die Implementierung ist jedoch komplexer und lässt sich nicht kurzfristig für alle Geräte umsetzen. Ähnliche Einschränkungen, aber auch unerwartete Potenziale, lassen sich auch für andere Muster als Erkenntnisgewinn identifizieren. In diesen Fällen handelt es sich also eher um Prototypen, die noch nicht ausreichend empirisch abgesichert sind. Die meisten der identifizierten Muster basieren aber auf existierenden Raumlösungen und Nutzungskonzepten. Dennoch wird erst die Zukunft zeigen, inwiefern sich der Raum etablieren wird. Die tatsächliche Raumnutzung wird also zu einer Bestätigung oder Falsifizierung der induktiv ermittelten Muster führen. Aus diesen Erfahrungen wird sich ableiten lassen, wo Muster weiter angepasst werden müssen, z.B. ob ein Kontext breiter oder enger gefasst werden muss. Auch die hypothetisch formulierten Wirkkräfte („Forces“), die eine Passung zwischen der beschriebenen Lösung und der Problemstellung begründen, sind so empirisch überprüfbar. Auf Basis dieser

Erkenntnisse sollen die Muster weiter ausgebaut und vollständig ausformuliert werden. Ziel ist die Entwicklung eines handlungsorientierten Booklets und Kartensets. Durch die gewonnenen Erfahrungen und weiterführende Forschung und Entwicklung fehlender Systemkomponenten, wird es in Zukunft leichter, weitere Räume in oder außerhalb der Hochschule entsprechend umzuformen oder neu einzurichten. Die Muster unterstützen dabei nicht nur den konkreten Gestaltungsprozess, sondern liefern Argumente, warum bestimmte Interventionen für ein kreativeres Arbeitsumfeld der Studierenden förderlich sind.

Literaturverzeichnis

- [Ad17] Adams Becker, S.; Cummins, M.; Davis, A.; Freeman, A.; Hall Giesinger, C.; Ananthanarayanan, V.: NMC Horizon Report: 2017 Higher Education Edition. The New Media Consortium, Austin, Texas, 2017.
- [Al77] Alexander, C.; Ishikawa, S.; Silverstein, M.; Jacobson, M.; Fiksdahl-King, I.; Angel, S.: A pattern language: towns, buildings, construction. Oxford University Press, New York, 1977.
- [Al79] Alexander, C.: The timeless way of building. Oxford University Press, New York, 1979.
- [Be14] Benyon, D.: Spaces of interaction, places for experience. Synthesis Lectures on Human-Centered Information 7/2, S. 1-129, 2014.
- [BM12] Benyon, D.; Mival, O.: Blended spaces for collaborative creativity. In: Proceedings of Workshop on Designing Collaborative Interactive Spaces AVI, 2012.
- [BM15] Benyon, D.; Mival, O.: Blended Spaces for Collaboration. Computer Supported Cooperative Work (CSCW) 24/2-3, S. 223-249, 2015.
- [BM16] Benyon, D.; Mival, O.: Designing Blended Spaces for Collaboration. Human Computer Confluence Transforming Human Experience Through Symbiotic Technologies 18, 2016.
- [Bo01] Borchers, J.: A pattern approach to interaction design. John Wiley & Sons, Hoboken, 2001.
- [Bu15] Bustamante, F. O.; Reyes, J. I. P.; Camargo, M.; DuPont, L.: Spaces to foster and sustain innovation: Towards a conceptual framework. In Engineering, Technology and Innovation/International Technology Management Conference (ICE/ITMC), 2015 IEEE International Conference, S 1-7, 2015.
- [Cr11] Cross, N.: Design thinking: understanding how designers think and work. Bloomsbury Academic, London, 2011.
- [DI13] DINI: Die Hochschule zum Lernraum entwickeln. Empfehlungen der Arbeitsgruppe „Lernräume“, University Press, Kassel, 2013.
- [FT08] Fauconnier, G.; Turner, M.: The way we think: Conceptual blending and the mind's hidden complexities. Basic Books, New York, 2008.

- [Fo14] Forest, C. R.; Moore, R. A.; Jariwala, A. S.; Fasse, B. B.; Linsey, J.; Newstetter, W.; Quintero, C.: The invention studio: A university maker space and culture. *Advances in Engineering Education* 4/2, S. 1-32, 2014.
- [Ga95] Gamma, E.; Helm, R.; Johnson, R.; Vlissides, J.: *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, Reading 1995.
- [Ho16] Hochschulforum Digitalisierung: *The Digital Turn – Hochschulbildung im digitalen Zeitalter*. Arbeitspapier 27, Hochschulforum Digitalisierung, Berlin, 2016.
- [JFW02] Johanson, B.; Fox, A.; Winograd, T.: The interactive workspaces project: Experiences with ubiquitous computing rooms. *IEEE pervasive computing* 1/2, S. 67-74, 2002.
- [KL01] Kelley, T.; Littman, J.: *The art of innovation: lessons in creativity from IDEO, America's leading design firm*. Currency, New York, 2001.
- [KP09] Kohls, C.; Panke, S.: Is that true...? Thoughts on the epistemology of patterns. In *Proceedings of the 16th Conference on Pattern Languages of Programs*. ACM, New York, 2009.
- [Ko14] Kohls, C.: The theories of design patterns and their practical implications exemplified for e-learning patterns, <https://opus4.kobv.de/opus4-ku-eichstaett/frontdoor/index/index/docId/158>. Stand: 28.03.2014.
- [KR14] Kristiansen, P.; Rasmussen, R.: *Building a Better Business Using the Lego Serious Play Method*. Wiley, Hoboken, 2014.
- [LNM09] Lindberg, T.; Noweski, C.; Meinel, C.: Design Thinking: Zur Entwicklung eines explorativen Forschungsansatzes zu einem überprofessionellen Modell. *Neuwerk Zeitschrift für Designwissenschaft* 1, S. 47-54, 2009.
- [MT16] Meinel, C.; von Thienen, J.: Design Thinking. *Informatik-Spektrum* 39/4, S. 310-314, 2016.
- [Mo07] Moultrie, J.; Nilsson, M.; Dissel, M.; Haner, U. E.; Janssen, S.; van der Lugt, R.: Innovation spaces: Towards a framework for understanding the role of the physical environment in innovation. *Creativity and Innovation Management* 16/1, S. 53-65, 2007.
- [PMW09] Plattner, H.; Meinel, C.; Weinberg, U.: *Design thinking: Innovation lernen - Ideenwelten öffnen*. Mi-Wirtschaftsbuch, München, 2009.
- [SL07] Schümmer, T.; Lukosch, S.: *Patterns for computer-mediated interaction*. Wiley series in software design patterns. John Wiley & Sons, Chichester, 2007.
- [SKB14] Skerlak, T.; Kaufmann, H.; Bachmann, G.: *Lernumgebungen an der Hochschule. Auf dem Weg zum Campus von morgen*. Waxmann, Münster, 2014.
- [St99] Streitz, N. A.; Geißler, J.; Holmer, T.; Konomi, S. I.; Müller-Tomfelde, C.; Reischl, W.; Steinmetz, R.: i-LAND: an interactive landscape for creativity and innovation. In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM, S. 120-127, 1999.
- [Ue15] Uebernickel, F.; Brenner, W.; Pukall, B.; Naef, T.; Schindlholzer, B.: *Design Thinking: Das Handbuch*. Frankfurter Allgemeine Buch, Frankfurt am Main, 2015.

PalmQA: Ein Frage-Antwort-Ensemble für selbstgesteuertes Lernen

Falk Pollok¹, Ulrik Schroeder¹, Mohamed Amine Chatti¹

Abstract: Eine große Herausforderung für selbstgesteuertes Lernen ist das selbständige Finden und Bewerten geeigneter Lernmaterialien. Zur Lernerunterstützung können Empfehlungssysteme genutzt werden, die die populärsten Materialien aufzeigen. Dabei werden weniger populäre Nischeninhalte per definitionem übersehen. Wir stellen ein Question-Answering-Ensemble vor, das als Ergänzung zu Empfehlungen und Filtern fungieren kann, indem es semantisches Verständnis nutzt, um Antworten zu natürlichsprachlich formulierten Fragen zu extrahieren. Durch die Kombination verschiedener quelloffener Systeme und Korpora konnte die Antwortrate gegenüber aktuellen vergleichbaren Einzelsystemen verdoppelt und die Antwortzeit um zwei Drittel verkürzt werden. Das System wurde in drei Nutzerstudien evaluiert, wobei sich sowohl breite Nutzerakzeptanz zeigte als auch im Quizwettbewerb zu Faktenfragen gegenüber 26 menschlichen Mitstreitern eine überlegene Leistung erzielt wurde. Wir diskutieren die Integration des Prototyps in Lernmanagementsysteme und stellen eine GUI-Erweiterung als Messaging-Dienst vor, die Themen interaktiv erschließbar macht.

Keywords: Question Answering, e-Learning, Cognitive Computing, Keyphrase Extraction, Distributional Semantics

1 Einleitung

Von Emanuel Goldbergs Statistischer Maschine und Vannevar Bushs Memex über Kollaborationssysteme wie Ted Nelsons Projekt Xanadu, Douglas Engelbarts oN-Line System (NLS) und Tim-Berners Lees ENQUIRE bis hin zu Apples Knowledge-Navigator-Konzept von 1987 ist die Geschichte der Informatik eng mit dem Thema von Wissenszugang verquickt, welches heute mit Persönlichen Assistenten wie Amazon Echo, Google Assistant, Microsoft Cortana und Apples Siri in greifbare Nähe gerückt ist. Eine neue Schlüsselrolle spielen dabei Question-Answering-Systeme (QA) wie IBMs DeepQA, das durch den Sieg gegen menschliche Grand Champions in Jeopardy 2011 breite Bekanntheit erlangt hat. Es gehört zum Paradigma des Cognitive Computing, in dessen Rahmen diverse Machine-Learning- und Neuro-Linguistische-Programmier-Techniken (NLP) genutzt werden, um menschlicher agierende Systeme zu entwickeln.

Die in diesem Artikel dargestellte Entwicklung, detailliert in [Pol16], zielt darauf ab, QA-Systeme für e-Learning und Forschung nutzbar zu machen. Sie knüpft u.a. an Piagets Beobachtung an, dass Kognition ein konstruktiver Prozess ist. QA-Systeme haben das

¹ Informatik 9, Learning Technologies, RWTH Aachen University Ahornstraße 55, 52074 Aachen, Germany, falk.pollok@rwth-aachen.de, [schroeder,chatti]@informatik.rwth-aachen.de

Potenzial, Lernen als individuelle Wissensschöpfung zu unterstützen, da sie Lernenden ermöglichen, ein Feld sowie die Zusammenhänge zwischen Konzepten für sich zu erkunden.

Ein Problem des selbstgesteuerten, informellen Lernens ist die eigenständige Auswahl der Lernmaterialien, weshalb Filter und Empfehlungssysteme als mögliche Unterstützung zum Einsatz kommen. QA-Systeme können eine dritte Säule in diesem Modell bilden, da sie große Textcorpora bearbeiten, relevante Passagen oder Antworten extrahieren und ihre Wahl semantisch begründen können. Ferner können sie den „Long Tail“ erschließen: Es gibt wenige sehr populäre Titel, den „Short Head“, sowie eine lange Reihe von Nischeninhalten, die durch Suchmaschinen und Mainstream-Medien kaum erschlossen werden. QA-Systeme sind nicht auf Popularitätsmetriken angewiesen, sondern agieren auf Basis natürlicher Sprachverarbeitung. Daher können sie helfen, sich durch Popularität hartnäckig haltende Fehlinterpretationen aufzulösen. Ferner sind QA-Systeme nur auf Text begrenzt, sondern funktionieren prinzipiell auch auf anderen Medien wie Bildern und Videos, was sie umso vielseitiger für den Bildungseinsatz macht.

1.1 Forschungsfragen

In der hier beschriebenen Arbeit wurde ein Prototyp eines Ensembles aus QA-Systemen für die Integration in Lernumgebungen entwickelt. Dabei mussten vor allem praktische Herausforderungen gelöst werden, um passable Antwortzeiten und -raten für die Unterstützung während Lernphasen zu erzielen und nachzuweisen, wie sich neue Domänen abdecken lassen. Auf dieser Basis sollten mögliche Mehrwerte von QA-Systemen für Forschung und Lehre erhoben werden. Mit dem Prototyp wurde untersucht, wie sich die Interaktion dynamisch gestalten und die Begrenzung auf Faktenfragen einschränken lässt.

2 Stand der Forschung

QA-Systeme erhalten eine Frage in natürlicher Sprache und ermitteln eine Antwort. Sie können entweder domänenoffen oder -spezifisch sein und sind häufig auf Faktenfragen beschränkt. Frühe Systeme wie Baseball [Gr61] sind nahezu ausschließlich domänenspezifisch. Natürlichsprachliche Datenbankschnittstellen wie LUNAR [Wo73] bieten den Vorteil, dass Nutzer Anfragesprachen und Struktur einer Datenbank nicht kennen müssen, um Anfragen zu stellen. In den 70ern kommen Wissensbasen und Expertensysteme auf, üblicherweise von Domänenexperten von Hand entwickelt, die eine weite Palette an Entscheidungsunterstützung abdecken. Während diese Zeit viele neue Konzepte wie Logikprogrammierung und regelbasierte Systeme hervorbringt, gelingt der Durchbruch nicht. Auch symbolische Ansätze zum Sprachverständnis bringen zwar Lösungen wie PoS-Tagging und Named Entity Recognition hervor, heben die Stagnation aber nicht auf, bis 1993 mit MURAX [Ku99] das Interesse wieder aufflammt, welches Information Retrieval mit natürlicher Sprachverarbeitung koppelt, was insbesondere im

Kontext von Mobilgeräten und Suchmaschinen Vorteile bieten kann. 1999 wird die Text Retrieval Conference TREC etabliert. TREC und das ACQUAINT-Programm bilden ein Fundament für IBM's PIQUANT [Pr03], auf welchem wiederum die spätere DeepQA-Architektur fußt, die mit Watsons Jeopardy-Sieg weite Aufmerksamkeit erfährt.

Charakteristisch für DeepQA [Fe10], welches auf der NLP-Architektur UIMA [Fe09] aufbaut, ist die Kombination von strukturierten und unstrukturierten Wissensquellen. Das zentrale Element ist eine so genannte Common Analysis Structure (CAS), welche das zu analysierende Artefakt ebenso beinhaltet wie alle Metadaten, die von Analyse-Komponenten generiert werden. Es existiert also eine gemeinsame Datenstruktur, in der alle Analyseergebnisse zentral erfasst werden können, so dass alle Komponenten auf den für sie relevanten Daten arbeiten können. Die Pipelines aus Lese-, Analyse- und Annotationsschritten können sowohl sequentiell als auch (via BPEL) nicht-sequentiell sein und verteilt sowie parallelisiert ausgeführt werden.

Um eine Antwort zu finden, extrahiert DeepQA zunächst Schlüsselphrasen aus der Frage sowie den Fokus - jenen Teil der Frage, der die Antwort repräsentiert und aus dem der Lexikalische Antworttyp (LAT) abgeleitet wird. Da keine Ontologie ausdrucksstark genug war, ist der LAT ein Bag of Words, also ein Multiset natürlicher Sprachausdrücke. Anschließend wird die Frage klassifiziert, was die nachfolgende Pipeline bestimmt. Zum Suchen von Antwortkandidaten werden drei Strategien verwendet: Title-Oriented Documents (TODs) wie Wikis, Textpassagensuche und strukturierte Daten. Für jeden Kandidaten wird die Quelle aus der er stammt hinterlegt. DeepQA nutzt nahezu 500 Scoring-Strategien, die via logistischer Regression in einem Konfidenzwert gebündelt werden, der mit der Antwort ausgegeben wird. Das offene System YodaQA [BŠ15] ist stark an diese Architektur angelehnt, beschränkt sich aber auf TODs, Fakten und Mainstream-Suchverfahren. Dennoch gehört es mit einer selbst angegebenen Akkuranz von 32.6% zu den besten offenen Systemen.

Im Rahmen des Educational QA (EQA) haben Wen et al. [We12] bereits QA-Systeme in LMS vorbereitet, sind allerdings in Antwortrate und -zeit durch ihr (Einzel-)Basissystem OpenEphyra limitiert. Gurevych et al. [Gu09] haben Community-QA-Seiten wie Yahoo Answers für Bildungszwecke nutzbar gemacht und sind im Ansatz, Fragen zu matchen, ähnlich zu Wang et al. [Wa06], die im Rahmen von Collaborative Learning FAQ-Kataloge abfragbar und dynamisch erweiterbar gemacht haben. Atapattu et al. [AFF15] haben EQA-Systeme erweitert, um automatisch Concept Maps zu generieren und so die strukturelle Bildung von Wissen zu unterstützen.

3 Innovationen

Der Prototyp PalmQA wurde als Microservice-Architektur (vgl. Abb. 1) realisiert, die verschiedene QA-Systeme einbindet, um die Antwortgenauigkeit für den Einsatz im eLearning deutlich zu steigern.

Das erste Ensemble bestand aus Amazon's Evi, Google, Kngine, MIT's Start, IBM's Watson QA API (eingestellt) sowie Wolfram Alpha. Später wurde auch Bings QA hinzugefügt. QAKiS wurde nicht mit aufgenommen, da es eine zu schlechte Präzision aufweist und seine Hauptquelle

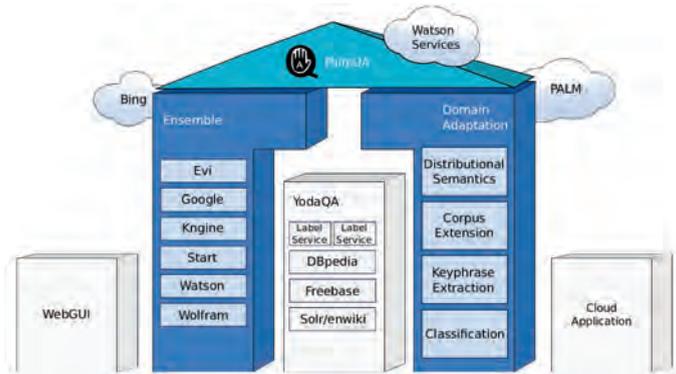


Abb. 1: Übersicht über die Microservice-Architektur

DBpedia bereits hinreichend durch YodaQA abgedeckt wird. Neben dem Ensemble wurden zwei graphische Nutzerschnittstellen für Inhaltsproduzenten und -konsumenten implementiert, ein Webinterface für die Integration in andere Systeme, ein Classifier, um vordefinierte FAQs abzufangen, eine neue Messaging-Metapher sowie Werkzeuge, um den Korpus zu erweitern, Schlüsselphrasen zu extrahieren und Distributional Semantics zu nutzen. Das Ensemble wurde erweitert, um neben QA auch auf generischen Backends wie den beiden letztgenannten zu funktionieren. Zudem wurden IBMs Clouddienste und PaaS Bluemix eingesetzt, was zu einem hybriden Clouddesign führte. Die Software kann dabei sowohl in der Cloud als auch on-premise ausgeführt werden, wobei für jeden Bluemixdienst eine On-Premise-Fallbacklösung bereitsteht.

3.1 Frage-Antwort-Ensemble

Das Ensemble unterstützt Question Answering, die Extraktion von Schlüsselphrasen sowie Distributional Semantics, alle in Dockercontainern gekapselt, sowohl für Antwortlisten (selektiert in Nutzerstudie) als auch Einzelaufgaben (quantitative Analyse).

Question Answering

Um zusätzliche QA-Backends hinzuzufügen, wurden APIs für Watson und Wolfram benutzt, wohingegen für alle anderen Quellen AJAX-Parsing via HTMLUnit zum Einsatz kam. Während der Pilotstudie wurden alle Nutzerinteraktionen zusammen mit den Systemantworten via Hibernate in eine H2-Datenbank geloggt, woraus eine Menge von 185 Anfragen hervorging, die dann benutzt wurde, um die Anzahl korrekter und inkorrektter Fragen sowie Enthaltungen zu ermitteln, aus denen schließlich Precision und Recall für jedes Backend abgeleitet werden konnte. Während das Ensemble komplexere Filter anbietet wie Thresholding von Konfidenzwerten oder Mischstrategien für Listen, etwa nach Mehrheitskriterien, liefert bereits die einfache Heuristik, Quellen nach absteigender Präzision auf dem Testdatensatz abzugehen und die erste gelieferte Antwort zu akzeptieren, sehr gute Ergebnisse. Diese Strategie hat zudem den Vorteil, Einzelantworten statt Listen zu liefern und nicht alle Quellen abfragen zu müssen.

Nachträgliches Ranking ist konzeptuell leicht integrierbar, derzeit aber nicht implementiert. Während kurz versucht wurde, Antwortkandidaten nach paarweisen Ähnlichkeitsmetriken (Block, Chapman, Cosine, Dice, Euclidean, Jaccard, Jaro, Jaro-Winkler, Levenshtein, Monge-Elkan, Needleman-Wunch, QGrams, Smith-Waterman, Soundex, Tag-Link) sowie dem längsten gemeinsamen Substring zu vereinen, war dies bisher nicht erfolgreich - vermutlich, da diese Strategie im Gegensatz zu Distributional Semantics semantische Ähnlichkeit außer Acht lässt. Da die mit der Heuristik erzielten Werte angenähert den optimalen entsprechen, wurde der Forschungsschwerpunkt zunächst auf andere Themen gelegt.

FAQ-Klassifikator Da viele Fragen zu komplex für aktuelle QA-Systeme sind und Lehrkräfte Antworten erweitern können sollen, werden häufige Fragen abgefangen und vordefiniert beantwortet. Um diese Funktionalität zu testen, wurden 100 Fragen aus zwei e-Learning-Vorlesungen sowie einem Seminar manuell extrahiert. Anschließend wurde jede Frage mehrfach paraphrasiert, um verschiedene Formulierungen im Datensatz abzubilden und dann einen Classifier darauf zu trainieren - online mit IBMs Natural Language Classifier, der auf CNNs basiert, und offline mit Random Forests sowie SVMs in Weka. Diese Umformulierung können in einem LMS durch Studierende oder z.B. via Crowdsourcing Plattformen wie Mechanical Turk erfolgen, um sowohl abgedeckte Themenanzahl als auch Genauigkeit zu verbessern. Für die eigene Klassifizierung wurden Fragen zunächst in Wortvektoren überführt und dann jedes Wort nach seiner Häufigkeit in der ersten Drittel Millionen von Häufigkeiten im Google Web Trillion Word Corpus sowie der Übereinstimmung mit Schlüsselphrasen aus dem Lehrstuhlkorpus gewichtet. Um Wekas Attribute-Relation File Format effektiv zu generieren, kam ein PATRICIA-Trie zum Einsatz, der es erlaubt, Worte nach Präfixen nachzuschlagen, wobei nur ein neuer Knoten und eine neue Kante hinzugefügt werden müssen, um einen neuen Eintrag abzubilden.

Korpus-Import Um Wissensbasis und Knowledge-Graphen erweitern zu können, müssen neue Dokumente importiert werden können. Der entsprechende Dienst in Bluemix ist allerdings nur bis 100MB kostenfrei. Während er sehr saubere Ausgaben liefert und sogar Formeln in ASCII-Repräsentationen überführt, wurde der Lehrstuhlkorpus mit mehreren Gigabyte an e-Learning-Artikeln lokal mit Apache Tika konvertiert. Zusätzlich wurde der LAK-Datensatz mit eigenem Code umgewandelt und Daniel Schneider von der Universität von Genf hat einen XML-Dump seines Edutech-Wikis zur Verfügung gestellt, welches nach manueller Säuberung in Solrs XML-Format gebracht wurde. Allerdings können die zugrundeliegenden Ontologien nicht auf domänenspezifische Anfragen angewendet werden und die Dokumentart wechselt von Wikis zu Textpassagen, was in einem so signifikanten Leistungseinbruch resultiert, dass dieses Thema für sich in zukünftigen Projekten erforscht werden muss.

Schlüsselphrasenextraktion

Die Schlüsselphrasenextraktion wurde sowohl mit Rapid Automatic Keyword Extraction (RAKE), welches hierfür auf Python 3 portiert wurde, als auch mit IBMs Alchemy API

durchgeführt. Das Ensemble kann je nach Wunsch der Anwendung beide Quellen parallel abfragen und nur von beiden gefundene Schlüsselphrasen akzeptieren, was einerseits die Qualität erhöht, andererseits aber auch gültige Einträge übersieht.

Distributional Semantics

Die durch den Wechsel von Jeopardy zu Fachdomänen beobachtete signifikante Leistungsverschlechterung wurde durch den Einsatz von Distributional Semantics durch das JoBim-Text-Framework [Bi13] realisiert. Der Ansatz folgt der Auffassung, dass die Bedeutung jedes Wortes vollständig durch die umgebenden Wörter abgeleitet werden kann. Dr. Martin Riedl (TUD) hat freundlicher Weise sein existierendes Wikipediamodell zur Verfügung gestellt und zusätzlich einen Distributional Thesaurus (DT) für den Lehrstuhlkorpus berechnet, da der lokale Aachener Hadoopcluster nicht genug Speicher dafür hatte. Mit diesem DT war es möglich, via JBT Worte zu clustern, Hearstmuster zu extrahieren und schließlich gelabelte Sinncluster zu berechnen, die in einer MariaDB-Instanz abgelegt in Echtzeit abgefragt werden können, was einen ersten Schritt hin zur Unterstützung domänenspezifischer Corpora darstellt.

Der praktische Nutzen ist, dass hiermit eine lexikalische Expansion durchgeführt werden kann, also etwa das Wort „exceptionally“ zu den sinnverwandten Begriffen „extremely, extraordinarily, incredibly, exceedingly, remarkably“ erweitert werden konnte, was für die erweiterte Betrachtung von LATs nützlich ist, wo jeder Kandidat in einem eigenen Thread betrachtet werden kann. Zusätzlich zu JBT verfügt das Ensemble nunmehr über die Word-Embedding-Ansätze word2vec und GloVe, um eine umfassende Grundlage für zukünftige Forschung zu bieten. Darüber hinaus können Word-Embeddings einfacher und ohne große Hadoopcluster berechnet werden.

Containerisierung

Um diese zahlreichen Microservices handhaben zu können, wurde jeder in ein eigenes Dockerimage gebündelt, welche gemeinsam mit Docker Compose orchestriert werden können. YodaQA etwa benötigt DBpedia, Freebase, enwiki in Solr sowie zwei Stichwortdienste und einen Container für sich selbst. Da Freebase und DBpedia auf derselben Technologie (Apache Fuseki) beruhen, können beide Container aus einem gemeinsamen Image gestartet werden. Dasselbe gilt für die Stichwortdienste. Welche Ordner als Volume gemountet werden sowie der genaue Befehl und die Umgebungsvariablen entscheiden, welchen „Phänotyp“ ein Image zur Laufzeit annimmt. Die Containerisierung wurde in das offizielle YodaQA-Repository aufgenommen.

Weitere Images sind das Ensemble, ein Backend, um alle webbasierten QA-Systeme anzusprechen, eines für Bluemix-Dienste, eines für die RAKE-Schlüsselphrasenextraktion sowie drei für Distributional Semantics: JBT zusammen mit seiner MariaDB-Datenbank in einem dedizierten Image sowie ein gemeinsames Abbild für beide Word-Embeddings-Ansätze. Um bei Erweiterungen Aufwand zu sparen, existieren generische Backends, die neben dem üblichen Kommunikationsweg über REST-Schnittstellen prinzipiell auch einen Thriftmodus anbieten, der etwa zum Einsatz kommt, um das Ensemble an die Plattform für Intelligente Assistenten Lucida anzubinden.

3.2 User Experience

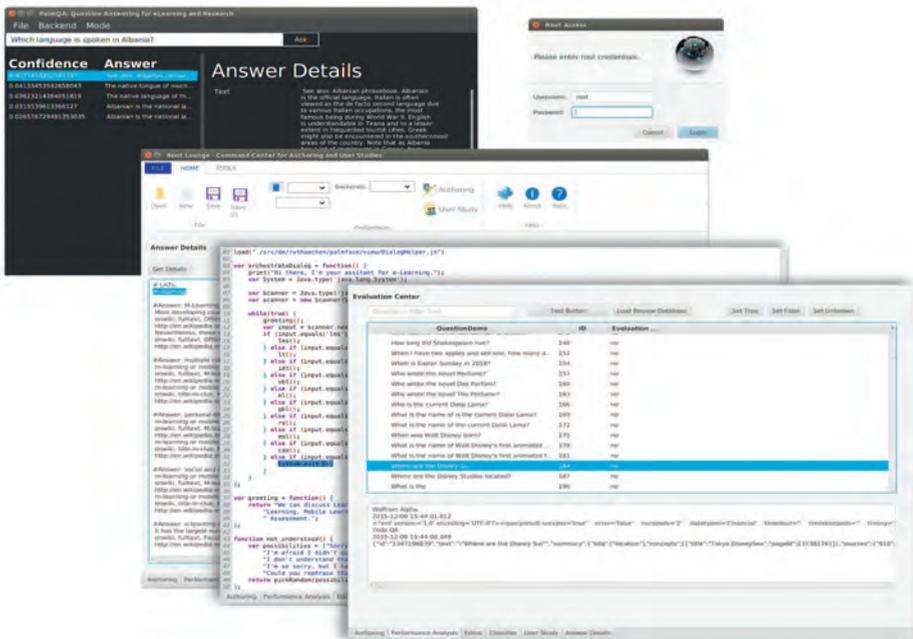


Abb. 2: Übersicht über die verschiedenen PalmQA-Benutzungsschnittstellen

Die GUI für Inhaltskonsumenten (Abb. 2) besteht aus vier Modi: Hauptmenü, QA, Dialogkomponente und eingebettetem Browser, um Zusatzquellen wie Wikipedia-Artikel, DBpedia-Einträge oder Bingtreffer sowie Wolframs Inhalte (insbesondere interaktive Pro-Features) anzuzeigen. Jene für Inhaltsproduzenten dient der Analyse von Antworten und Zusatzaufgaben wie dem Skripten von Dialogsequenzen.

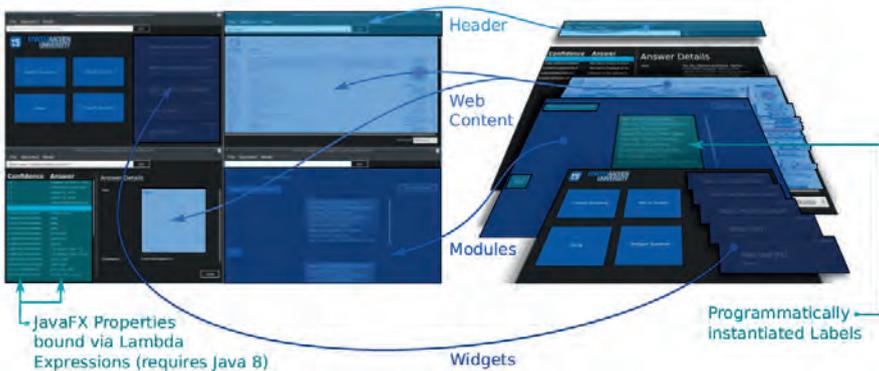


Abb. 3: GUI Modularisierung

Genauso wie die Backends als entkoppelte Microservices abgebildet wurden, ist die Benutzeroberfläche ebenfalls hochgradig modular (vgl. Abb. 3): Die Kopfzeile ist unabhängig von den auswechselbaren Hauptmodulen, sowohl Antworten als auch ihre Belege werden in WebViews eingebettet dargestellt, Listen sind JavaFX-Properties, die durch Lambda-Ausdrücke an das entsprechende Feld gebunden wurden und Dialogfelder ebenso wie FAQ-Widgets auf der Menüseite werden programmatisch generiert. Jede GUI-Komponente ist in FXML definiert und verfügt über einen eigenen Controller, wobei komplexe Interaktionen und Kommunikation untereinander über einen Hauptcontroller abgewickelt werden.

Messaging-Metaphorik

Um die Interaktion gegenüber herkömmlichen QA-Systemen dynamischer zu gestalten, wurde als Nebenprojekt eine Messaging-Metaphorik entwickelt (Abb. 4), die einerseits QA-Systeme in einer an Instant-Messaging angelehnten Kontaktliste darstellt und andererseits einfache, geskripte Dialoge anbietet, um Themen interaktiv erkunden zu können.

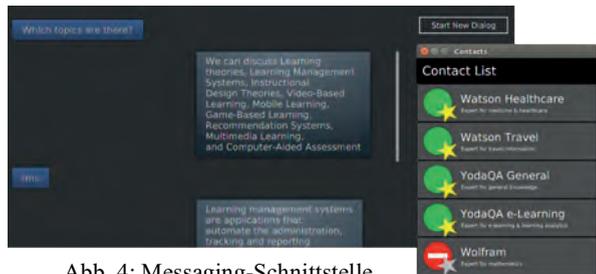


Abb. 4: Messaging-Schnittstelle

Hierzu werden im Grußtext Oberthemen genannt, die dann mit Folgefragen erschlossen werden können. Diese Metapher ist keine Kernkomponente und daher nicht als vollständiges Dialogsystem angelegt, sondern dient als Demonstration, um in den Nutzerstudien ein Vorabbild zukünftiger Erweiterungen erfragen zu können.

Web-GUI und Integration

Das auf der Adressbuchdemo von Vaadin basierende Webinterface, kann als iFrame eingebunden werden. Auf dieser Grundlage wurde das System mit der RWTH Lehr- und Lernplattform L²P integriert. In der Nachfolge erfolgte die Integration mit der Plattform für Intelligente Assistenten Lucida, die an der University of Michigan entwickelt wird, wodurch das mitgelieferte QA-System OpenEphyra durch das in Geschwindigkeit und Antwortrate überlegene PalmQA-Ensemble ersetzt werden kann. Das Projekt hat das Ensemble mit in sein Repository aufgenommen.

4 Evaluation

4.1 Quantitative Evaluation

Zunächst sei die Antwortgeschwindigkeit auf den später detaillierten „Grand-Challenge-Fragen“, welche viele Fragekategorien abdecken, grob analysiert: YodaQA war in der

Standardkonfiguration mit nicht-lokalen Datenquellen am langsamsten (Durchschnitt 23.9s, Beschleunigungsfaktor 1, [min. 5s .. max. 64s]), etwas schneller in der offiziellen Demo auf ailao.eu (19.8, 1.21, 2..51) und bereits doppelt so schnell, wenn die Datenquellen ebenfalls lokal ausgeführt werden (12.03, 1.98). YodaQA zeigte dabei auffällig hohe IO-Raten, was darauf schließen lässt, dass IO eher Flaschenhals als reine CPU-Kraft ist. Diese Beobachtung bestätigt sich beim Einsatz einer einfachen Consumer-SSD (8.53, 2.8, 1..25), die nahezu dreifache Beschleunigung liefert. Interessanter Weise verbessern weder größere SSDs noch eine professionelle Samsung 950 Pro mit M.2 NVMe über PCIe 3.0 mit deutlich besserem Datendurchsatz (1582MB/s statt 269MB/s für vorige 850 Evo) diesen Wert erheblich. Selbst mit Freebase auf der 850 Evo und allen anderen Backends auf der 950 Pro bleibt die Geschwindigkeit nahezu stabil (8.3, 2.88, 1..19) mit leichter Verbesserung für den ungünstigsten Fall. Die Leistungseinbußen durch Dockercontainer sind mit etwa 7% vertretbar (9.53, 2.51, 2..28) vs. (10.2, 2.34, 2..29) mit allen Backends auf 850 Evo SSD). Bei einer Extremfallmessung mit jeder Quelle auf einer dedizierten SSD hat sich die Geschwindigkeit sogar im Vergleich zu nur einer SSD verschlechtert (10.2, 2.34, 1..29). Auch Arbeitsspeicher mit 32GB gegenüber 16GB (8.6, 2.78, 1..27) zu (9.23, 2.59, 2..26) hat einen messbaren, aber nicht erheblichen Einfluss, weshalb festzuhalten ist, dass einzig die Nutzung einer lokalen SSD eine maßgebliche Verbesserung einbringt, wohingegen alle anderen Parameter verblassen.

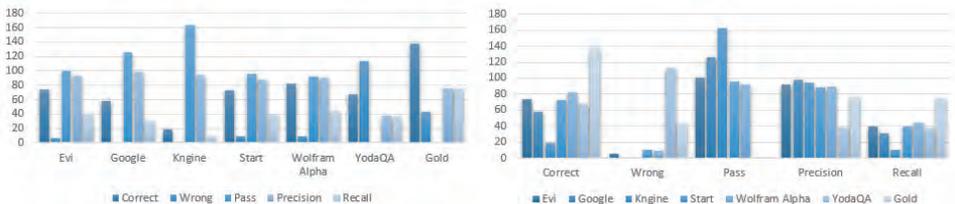


Abb. 5: Übersicht über Quantitative Evaluation

Um einen Eindruck der Ensemblekomponenten zu erhalten, wurden diese gegen die 185 während der Nutzerstudien gesammelten faktoiden Fragen getestet (vgl. Abb. 5). Hieraus ergibt sich die folgende Ordnung nach Präzision, wobei die Tupel die Anzahl korrekter Antworten, dann teilkorrekt beantworteter Antworten, inkorrekt beantworteter Antworten, Enthaltungen sowie Präzision und Recall wiedergeben: Google (58, 0, 1, 126, 0.98, 0.31), Kngine (19, 1, 1, 163, 0.95, 0.10), Evi (74, 5, 6, 100, 0.925, 0.4), Wolfram Alpha (82, 2, 9, 92, 0.90, 0.44) und Start (73, 5, 10, 96, 0.88, 0.39). Wie sich zeigt, weisen alle Backends hohe Präzision auf, was eine Zusammenführung von Antworten aus unterschiedlichen Quellen begünstigt.

4.2 Nutzerstudien

Die Kernevaluation erfolgte in drei Nutzerstudien - einer Pilotstudie, einer Hauptstudie und einer „Grand Challenge“. Die Motivation für mehrere kleine Studien gegenüber einer groß angelegten findet sich in mehreren Studien zur idealen Probandenanzahl, etwa in [DM90] oder einer Zusammenfassung von 16 Studien durch Nielsen und Landauer

[NL93], in der festgestellt wurde, dass das ideale Kosten-Nutzen-Verhältnis bereits mit 4 Teilnehmern erreicht wird und eine Sättigung bei etwa 16 Subjekten eintritt. In gewisser Weise tritt das Divide-and-Conquer-Prinzip von Softwarearchitektur und Benutzeroberfläche somit auch beim Studienentwurf in Erscheinung.

Die Pilotstudie folgt einem Within-Subjects-Design mit 12 Teilnehmern, die eine frühe Version des Ensembles mit Watson verglichen und dann zu ihrem allgemeinen Eindruck von QA-Systemen befragt wurden. Alle Alters- und Bildungsstufen außer „Bachelor at University of Applied Science“ wurden abgedeckt ebenso wie alle Sprachfertigungsstufen in Englisch. Das Interesse war auffällig hoch mit 812 gestellten Fragen, was 67.7 Fragen pro durchschnittlichem Nutzer entspricht. Während die kürzeste Bearbeitungszeit 24 Minuten betrug, verbrachten einige Nutzer freiwillig über 2 Stunden damit, mit dem System zu experimentieren. Das frühe Ensemble wurde durchgängig als vielversprechend (7) oder bereits nützlich (5) bewertet, wohingegen Watson konsistent als schneller und von 8 als bereits nützlich, von 3 Nutzern aber auch als noch viel Arbeit benötigend eingeschätzt wurde. Watson wurde dabei vor allem als nützlich für Erklärungen charakterisiert (5 vs. 1, sowohl für den Medizin- als auch Reisekorpus). Unerwartet wurde jedes der beiden Systeme von 5 Probanden als Unterstützung bei der Forschung in Erwägung gezogen, wohingegen niemand die Systeme als Spielzeug oder technische Revolution ansah. 8 beschrieben die Systeme als wertvolle Ergänzung zu Suchmaschinen und drei sahen sogar das Potenzial, Suchmaschinen zu ersetzen. Insgesamt sahen 11 Nutzer die Systeme als grundsätzlich hilfreich in der Lehre und 7 in der Forschung an.

Die Hauptstudie verglich YodaQA ohne Modifikationen mit dem vollen PalmQA Ensemble, basierte diesmal auf einem Between-Subjects-Design und wurde in einem lateinischen 2×2 Quadrat systematisch variiert, um auszugleichen, welches System zuerst benutzt wurde und damit als Vergleichskandidat fungierte. Die 16 Teilnehmer waren gleich auf die Geschlechter unterteilt. Auf einer 5-stufigen Likert-Skala bewerteten die Probanden das PalmQA Ensemble klar als schneller (med. 4.5, avg. 4.44) und berichteten nahezu übereinstimmend, dass es mehr korrekte Antworten liefere (5, 4.81). Ebenso wurden Antwortqualität (4, 4.31), Antwortdetails (4, 4.06) und die Nachrichtenmetapher (4, 4.19) positiv bewertet, die vordefinierten Antworten wurden sogar als exzellente Ergänzung eingeschätzt (5, 4.625).

Die finale Nutzerstudie wurde als „Grand Challenge“ an IBMs Jeopardy-Herausforderung angelehnt, in der 26 Probanden (13 männlich, 13 weiblich) bei 30 Faktenfragen, die ein nicht in das Projekt involvierter Kollege frei verfasst hat, gegen das System antraten. Der bestplatzierte Mensch schaffte es, 24 Fragen korrekt zu beantworten, was deutlich über dem Durchschnitt von 11 Fragen lag. Dennoch war das PalmQA Ensemble in der Lage, mit 25 Fragen knapp zu gewinnen und mehr als doppelt so viele Antworten wie der Durchschnitt der Probanden zu liefern, was statistisch hochgradig signifikant ist: Ein einseitiger t-Test (single sample) zum Signifikanzniveau 0.01 liefert einen p-Wert unter 0.00001.

5 Ergebnisse

Während das PalmQA Ensemble mit mehr als doppelt so vielen korrekten Antworten wie das beste offene Einzelsystem eine „übermenschliche“ Antwortrate aufweist mit durchschnittlicher Antwortzeit unter 10 Sekunden, was einer Minderung um fast einen Faktor von 3 entspricht, hat es auch klare Grenzen aufgezeigt. Es ist im Stande, etwa drei von vier Fragen richtig zu beantworten und dabei Quellen wie Wikipediaartikel zu liefern, was einen Startpunkt für weitere Recherchen in Forschung und Lehre bietet, verfügt aber über kein tiefes Verständnis, um ein Feld umreißen oder eigene Erklärungen formulieren zu können oder selbst sinnvolle Zusatzinformationen anzumerken, wo es hilfreich scheint. Aufgezeigte Erweiterungen wie das Abfangen vordefinierter Fragen können diese Unzulänglichkeiten kaschieren und somit die Akzeptanzrate steigern, sind aber nur temporäre Hilfslösungen. Auch die Nachrichtenmetapher verbessert das Basissystem, da sie ein Thema interaktiv erkundbar macht. Als Fazit scheint es für die Lehre ratsam, Lehrpersonal aktiv in das Antwortsystem einzubeziehen. Die größte Herausforderung im Hinblick auf die Zielsetzung dieser Arbeit bleibt jedoch die Forschung. Während momentane Antwortraten für Fragen zu Allgemeinwissen hinreichend sind, muss ein System für die Forschung vor allem domänenspezifisches Wissen zugänglich machen. Während Distributional Semantics, Keyphrase Extraction und Dokumentenkonvertierung erste Schritte in diese Richtung bieten, bleibt ein langer Weg, insbesondere, da zugrundeliegende Systeme wie YodaQA das Format von Title-Oriented-Documents oder die Verfügbarkeit von Ontologien annehmen, was die Domänenanpassung erschwert.

Die Eigenschaft, dass das PalmQA Ensemble sowohl Listen- als auch Einzelantworten formen kann, wurde bisher nicht ausgeschöpft. Dabei eröffnet es die Möglichkeit, auf die Zielgruppe zugeschnitten zu antworten. Nach der Erfahrung mit Watson-QA-APIs ist es für Forscher deutlich nützlicher, Textpassagen statt isolierter Antworten auszugeben und sie sind auch eher bereit, eine Liste von Antwortkandidaten durchzugehen.

Summa summarum bietet das System eine neuartige Plattform, um QA-Systeme in Forschung und Lehre nutzbar zu machen. Mit technischen Fortschritten im Cognitive Computing allgemein und in Dialogsystemen im Speziellen kann dieser Grundstock sukzessive erweitert werden, bis das Ziel vom intelligenten Assistenten für Fachwissen erreicht ist.

Literaturverzeichnis

- [AFF15] Atapattu, Thushari, Katrina Falkner, and Nickolas Falkner: Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, In . C. Conati, N. Heffernan, A. Mitrovic, and F.M. Verdejo, eds. Pp. 13–22. Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-19773-9_2, 2015.
- [BR13] Biemann, C., Riedl, M.: Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity. *Journal of Language Modelling* 1(1): 55–95, 2013.

- [BŠ15] Baudiš, Petr; Šedivý, Jan: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 6th Int. Conference of the CLEF Association, CLEF'15, Toulouse, France, September 8-11, 2015, Proceedings. Springer International Publishing, Cham, Kapitel Modeling of the Question Answering Task in the YodaQA System, S. 222–228, 2015.
- [DM90] Dalal, S. R.; C. L. Mallows: Some Graphical Aids for Deciding When to Stop Testing Software. *IEEE Journal on Selected Areas in Communications* 8(2): 169–175, 1990.
- [Fe09] Ferrucci, David, Adam Lally, Karin Verspoor, and Eric Nyberg: Unstructured Information Management Architecture (UIMA) Version 1.0, OASIS Standard, 2009.
- [Fe10] Ferrucci, David A., Eric W. Brown, Jennifer Chu-Carroll, et al.: Building Watson: An Overview of the DeepQA Project. *AI Magazine* 31(3): 59–79, 2010.
- [Gr61] Green, Bert F., Jr., Alice K. Wolf, Carol Chomsky, and Kenneth Laughery: Baseball: An Automatic Question-Answerer. Western Joint IRE-AIEE-ACM Computer Conference Pp. 219–224. IRE-AIEE-ACM '61 (Western). New York, NY, USA: ACM. <http://doi.acm.org/10.1145/1460690.1460714>, 1961.
- [Gu09] Gurevych, Iryna, Delphine Bernhard, Kateryna Ignatova, and Cigdem Toprak: Educational Question Answering Based on Social Media Content. In Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling Pp. 133–140. Amsterdam: IOS Press. <http://dl.acm.org/citation.cfm?id=1659450.1659474>, 2009.
- [Ku99] Kupiec, Julian M.: Natural Language Information Retrieval. In . Tomek Strzalkowski, ed. Pp. 311–332. Dordrecht: Springer Netherlands, 1999.
- [NL93] Nielsen, Jakob, and Thomas K. Landauer: A Mathematical Model of the Finding of Usability Problems. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems Pp. 206–213. CHI '93. New York, NY, USA: ACM. <http://doi.acm.org/10.1145/169059.169166>, 1993.
- [Pol16] Pollok, Falk: Leveraging UIMA-Based Question Answering Systems for e-Learning and Research. Master's Thesis, RWTH Aachen, Learning Technologies, 2016.
- [Pr03] Prager, John M.; Chu-Carroll, Jennifer; Czuba, Krzysztof; Welty, Christopher A.; Ittycheriah, Abraham; Mahindru, Ruchi: IBM's PIQUANT in TREC2003. In (Voorhees, Ellen M.; Buckland, Lori P., Hrsg.): TREC. Jgg. Special Publication 500-255. National Institute of Standards and Technology (NIST), S. 283–292, 2003.
- [Wa06] Wang, Chun-Chia, Jason C. Hung, Che-Yu Yang, and Hsuan-Pu Chang: A Question Answering System Approach for Collaborative Learning. In CSCWD 2006, Southeast University, Nanjing, China Pp. 1403–1407.
- [We12] Wen, Dunwei, John Cuzzola, Lorna Brown, and Dr Kinshuk: Instructor-Aided Asynchronous Question Answering System for Online Education and Distance Learning. *The International Review of Research in Open and Distributed Learning* 13(5): 102–125, 2012.
- [Wo73] Woods, W. A.: Progress in Natural Language Understanding: An Application to Lunar Geology. In: Proceedings of the June 4-8, 1973, National Computer Conference and Exposition. AFIPS '73, ACM, New York, NY, USA, S. 441–450, 1973.

Entwicklung und Durchführung computerbasierter Tests zur Messung von Musikkompetenzen

Julia Finken¹, Franziska Marx², Michaela Meyer³, Philipp Krieter⁴ und Andreas Breiter⁵

Abstract: Das Fach Musik stellt besondere Herausforderungen an die Kompetenzmessung. Das gilt sowohl für die psychometrischen Modelle als auch für die Entwicklung computergestützter Testsysteme. Insbesondere die Verwendung von Bild- und Tonbeispielen als auch die Messung von Rhythmus- und Melodieverständnis setzen neuartige multimediale Testitems voraus. In Zusammenarbeit von Musikpädagoginnen und -pädagogen und Informatikerinnen und Informatikern wurde im Projekt „PosyMus – Potenziale von Feedbacksystemen im Musikunterricht“⁶ ein neuartiges Testsystem entwickelt und in Schulen getestet. Der Beitrag stellt die Testitems sowie die Erhebungen in den Schulen vor. Dabei wurde zweierlei deutlich: Einerseits fördern die computerbasierten Aufgabentypen die Motivation von Lernenden und andererseits sind zahlreiche praktische Herausforderungen zur Einbettung in den Unterrichtsalltag zu lösen. Insbesondere die Rückmeldung der Testergebnisse wurde als Anforderung formuliert. Diese wird in Zukunft im Projektverlauf durch ein elektronisches Rückmeldesystem umgesetzt.

Keywords: Musikunterricht, Computerbasiertes Testen in Schulen, Kompetenzmessung, Multimediale Testitems.

1 Einleitung

Die Messung von Kompetenzen hat in der empirischen Bildungsforschung in der letzten Dekade erheblich an Bedeutung gewonnen. Kompetenzen werden hierbei verstanden als „die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven 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“ [We01]. Die hierzu erforderlichen standardisierten Tests wurden zunächst für die Kernfächer entwickelt und zunehmend auch mit Hilfe computerbasierter Testsysteme umgesetzt. In einem weiteren Schritt haben Studien in der Musikpädagogik gezeigt [Ni08], [JKL10], [Jo12], [HL14], dass Kompetenzmessungen auch in anderen Fächern möglich sind. Eine besondere Herausforderung im Fach Musik stellt die Vielfalt des Kompetenzerwerbs dar, die in erster Linie auf kreativ-ästhetische Facetten verweist. Aus psychometrischer Sicht sind Rhythmuserkennen und Hörverstehen neben klassischen

¹ Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, jfinken@ifib.de

² Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, fmarx@ifib.de

³ Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, mmeyer@ifib.de

⁴ Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, pkrieter@ifib.de

⁵ Institut für Informationsmanagement Bremen GmbH, Am Fallturm 1, 28359 Bremen, abreiter@ifib.de

⁶ gefördert vom Bundesministerium für Bildung und Forschung, Förderkennnummer 01LSA1506A/B

Wissensfragen zur Harmonielehre oder Musikgeschichte relevant und umsetzbar. Hierfür eignen sich computergestützte Verfahren deshalb besonders gut, weil Hörbeispiele bzw. das Nachspielen von Melodien so auf vielfältige Weise und anhand einer großen Anzahl von Beispielen ermöglicht werden können. Bestimmte Teilbereiche können im klassischen Unterricht bisher nicht individualisiert erfasst und gespeichert werden. Computerbasiertes Testen (CBT) ermöglicht, durch die Einbettung in elektronische Rückmeldesysteme, die detaillierte Betrachtung von Individual- oder Gruppenleistungen. In diesem Beitrag stellen wir einen Ansatz für ein computerbasiertes Testsystem im Bereich Musik sowie die Durchführung der Tests in Schulen vor. Bei der Zielgruppe handelte es sich um Schülerinnen und Schüler aus weiterführenden Schulen (Sekundarstufe I) ohne Vorgaben oder Einschränkungen bezüglich der musikalischen Fertigkeiten und Vorkenntnisse.

2 Stand der Forschung

2.1 Computerbasiertes Testen

Jurecka und Hartig definieren CBT als „die Verwendung von Computern zur Erfassung psychologischer Merkmale und Konstrukte. Hierbei erfolgt zumindest die Vorgabe des Testmaterials am Computerbildschirm [...] und die getestete Person reagiert ebenfalls über einen Computer“ [JH07]. Die Nutzung computergestützter Verfahren zur Testdurchführung ist sowohl im Bereich der Hochschulbildung (z. B. eKlausuren) als auch in der beruflichen Bildung (z. B. Handelskammer-Prüfungen) sowie bei der Überprüfung von Sprachkompetenzen (z. B. TOEFL) und anderen Prüfungen (z. B. Führerschein) bereits umfangreich erforscht und erprobt worden. Auch in der empirischen Schulforschung wurden in den letzten Jahren computer- anstatt papierbasierte Tests eingesetzt, z. B. für Sprachstandserhebungen wie CITO [SK08] oder bei PISA (z. B. [Ze15], [Ma09]). Als Vorteile von CBT für die Kompetenzmessung nennen Jude und Wirth [JW07] die Möglichkeit zur Erfassung komplexer und dynamischer Problemlösefähigkeit, die Darbietung multimedialer Aufgaben- und Testformate sowie die Erfassung von Verhaltensbereichen, die über die Möglichkeiten von Papier und Stift hinausgehen und benennen dabei explizit die Musik als Anwendungsbereich (z. B. Hörbeispiele zur Identifikation von Akkordstrukturen [JW07]). Die Erwartung liegt darin, bestehende Kompetenzen besser testen zu können und weitere Kompetenzen überprüfbar zu machen. Zudem eröffnen computergestützte Verfahren die Chance zum „adaptiven Testen“ [GI00], d. h. zur Anpassung der Testaufgaben an das jeweilige Niveau der Getesteten.

Für die Entwicklung von Aufgaben für CBT stehen mittlerweile zahlreiche Werkzeuge zur Verfügung. Das PISA-Konsortium arbeitet mit TAO und dem ItemBuilder⁷. In den USA gibt es einen starken Wettbewerb über die Bildungsverlage, während in Deutsch-

⁷ <http://tba.dipf.de/de/assessment/software>

land einige Anbieter im beruflichen Bereich sowie im Hochschulbereich tätig sind. Speziell für die Schule hat sich jedoch noch kein Markt entwickelt. Die Schulbuchverlage produzieren teilweise eigene Software, deren Wiederverwendbarkeit aufgrund der fehlenden Interoperabilität nahezu unmöglich ist. Die Bundesländer haben sich z. B. über die Kultusministerkonferenz bisher nicht auf einheitliche Datenstandards einigen können. Keine Beachtung hat bisher die Interoperabilität innerhalb der gesamten Prozesskette gefunden, die von der Generierung von Testitems und ihrer Erschließung über Aufgaben-Datenbanken, die Durchführung und Ergebnisse der Tests bis zu ihrer Verwendung in computergestützten Rückmeldesystemen reicht [St12].

2.2 Computerbasiertes Testen im Bereich Musik

International beschäftigt sich der Arbeitskreis um Vispoel mit computerbasiertem Testen im Bereich Musik. So konnte gezeigt werden, dass adaptives computerbasiertes Testen geeignet ist, um Hörfähigkeit zu messen [VWB97]. Auch die Testungen im Bereich Music Assessment im Rahmen des South Carolina Arts Assessment Program (SCAAP)⁸ werden seit 2000 computerbasiert durchgeführt [Ya08]. Neben dem Test wird die Internetplattform auch genutzt, um eine Beurteilung der offenen Schülerinnen- und Schülerantworten vorzunehmen. Weiterhin findet CBT Verwendung bei der Erfassung von rhythmischen Fähigkeiten [As11] unter Einsatz des Programms SmartMusic. Balizet et al. [BTP99] haben darüber hinaus papier- und computerbasierte Verfahren im Hinblick auf die Messung der Hörfähigkeit verglichen und sind, wie auch die Forscher im KoMus-Projekt [JKL10], zu dem Schluss gekommen, dass sich insbesondere im Bereich der Hörfähigkeit dem computerbasierten Test viele Vorteile (z. B. die Möglichkeit des individualisierten Hörens) zuschreiben lassen. Denen stehen einige Schwierigkeiten gegenüber, die sich z. B. aus der Einbindung von Hörbeispielen oder Noten ergeben. Allerdings hat die Entwicklung von Anwendungen auf der Basis von mobilen Endgeräten hier neue Möglichkeiten eröffnet, die auch für PosyMus relevant sind. So haben Hasselhorn und Lehmann [HL14] Tablets mit einer spezifischen anschlussensiblen Programmierung als Tastatur farbasiert in einer notenfreien Umgebung verwendet. Zudem existieren sowohl offene als auch kostengünstige Apps, die vielfältige Tonhöhen- und Rhythmeingaben ermöglichen [Do13].

3 Das Testsystem

3.1 Anforderungen an das Testsystem

Die Herausforderungen eines computerbasierten Testsystems für den Bereich Musik brachten einige Neuerungen mit sich, die schon vor der Entwicklung des Systems beachtet werden mussten. Zunächst war es notwendig, die Anforderungen an das System

⁸ <https://scaap.ed.sc.edu/>

zu bestimmen. Dabei standen nicht nur technische, sondern vor allem auch musikalische bzw. musikpädagogische Anforderungen im Fokus [FMB16]. Aus diesem Grund wurden die später zu implementierenden Testitems sowie die Art der Nutzung durch die Testteilnehmenden erfasst, um daraus Design und Softwarearchitektur abzuleiten. Diese Schritte wurden während des Entwicklungsprozesses wiederholt, um die Anforderungen an das Testsystem stetig zu präzisieren [GH06]. Durch dieses User-Centered Design werden die künftigen Testteilnehmenden mit ihren Eigenschaften, Zielen und Aufgaben in den Mittelpunkt des Entwicklungsprozesses gestellt [AMP04], [Ni93].

3.2 Aufbau des Testsystems

Das Testsystem besteht aus mehreren Soft- sowie Hardwarekomponenten. Der Test wird mithilfe einer Software (TAO [FMB16]) auf einem Testserver zur Verfügung gestellt. Die Schulen stellen Tablets (entweder schul- oder schülereigene) sowie Kopfhörer zur Verfügung. Um einen etwaigen Mangel dieser auszugleichen, ist es sinnvoll, dass die Testleitung sowohl Tablets als auch Kopfhörer als Reserve mitbringt.

Im Projekt wurde sich für die Verwendung von zwei verschiedenen Serverlösungen zur Bereitstellung des Testsystems entschieden. Beide Testsysteme sind dabei identisch eingerichtet, so dass sie gleichwertig verwendet werden können. Die erste Lösung liegt auf einem virtuellen Server und ist über das Internet erreichbar. Um diese nutzen zu können, muss sichergestellt sein, dass die Internet- sowie Datenverbindung in der Schule ausreichend und stabil genug ist, damit eine große Anzahl von Schülerinnen und Schülern gleichzeitig auf das Testsystem zugreifen kann, ggf. parallel zur Nutzung durch andere Klassen. Ebenfalls zu beachten ist die WLAN-Abdeckung der Schule, da der Empfang nicht in allen Räumen gegeben ist. Da nicht alle Schulen diese Voraussetzungen erfüllen, wurde ein zweites Testsystem auf einem lokalen, transportablen Server eingerichtet. Mithilfe eines ebenfalls transportablen Routers wird ein lokales Netzwerk zu diesem Server aufgebaut. Die Schülerinnen und Schüler können sich über die WLAN-Einstellungen ihres Tablets mit diesem Netzwerk verbinden und so auf den Test zugreifen. Dieses Vorgehen stellt gleichzeitig sicher, dass keine anderen Anwendungen im Internet genutzt werden können, so dass die Schülerinnen und Schüler hierdurch weder abgelenkt werden, noch ggf. nach Lösungen für die Aufgaben im Internet suchen oder sich über die Aufgaben austauschen können.

Zur Durchführung des Tests benötigen alle Teilnehmenden ein WLAN-fähiges Tablet sowie Kopfhörer. Der Test ist webbasiert und wird über einen Webbrowser aufgerufen und durchgeführt. Die Anwendung ist für Chrome sowie Safari optimiert, da diese gängige Browser auf Tablets sind. Innerhalb der Testitems werden Musikstücke verwendet, die zum einen individuell abgespielt, zum anderen nachgespielt werden müssen. Durch die Kopfhörer wird der Lärmpegel gesenkt und sichergestellt, dass die Teilnehmenden ungestört die Aufgaben wahrnehmen und bearbeiten können.

3.3 Testitems

Für das Projekt wurden in Anlehnung an Hasselhorn [Ha15] Aufgabentypen aus den Bereichen „Musik wahrnehmen und kontextualisieren“ sowie „Musikpraxis“ ausgewählt. Die Aufgaben wurden im Rahmen der Forschungsprojekte KoMus⁹ und KOPRA-M¹⁰ entwickelt und validiert [JKL10]. Der Bereich „Musik wahrnehmen und kontextualisieren“ enthält Items, bei denen bspw. die Anzahl der Abschnitte eines Musikstücks genannt werden sollen oder ein Instrument einem vorgespielten Musikstück zugeordnet werden muss. Bei den meisten Aufgaben werden den Schülerinnen und Schülern verschiedene Antwortmöglichkeiten angeboten, aus denen sie die richtige Antwort auswählen sollen. Die Fragen und Antworten beziehen sich hierbei auf kurze Musikstücke, die die Schülerinnen und Schüler sich vor Auswahl der Antwort anhören können (vgl. Abb.1).

Wie viele Abschnitte hörst Du?

Du hörst gleich ein kurzes Musikstück. Wie viele Abschnitte hat es?

Abspielen

Zwei Abschnitte Drei Abschnitte Vier Abschnitte Fünf Abschnitte

Speichern und weiter

Abb.1: Wie viele Abschnitte hörst Du?

Zusätzlich gibt es Aufgaben, bei denen das freie Assoziieren gefördert wird. Hierbei sollen die Schülerinnen und Schüler sich z. B. vorstellen, dass sie einen Artikel über einen Bandwettbewerb schreiben sollen und sich hierfür Kriterien überlegen und stichpunktartig notieren, auf die sie beim Anhören der Bands achten wollen.

Der Bereich „Musikpraxis“ unterteilt sich nach Hasselhorn [Ha15] in drei Bereiche: Rhythmusproduktion, Melodieproduktion und Gesang. Da der Bereich „Gesang“ erweiterte Anforderungen an den Aufbau der Testumgebung hat (auditive Abschirmung der Teilnehmenden), wurde hierauf im Rahmen von PosyMus verzichtet. Die Items aus den Bereichen „Rhythmusproduktion“ und „Melodieproduktion“ sind ähnlich aufgebaut. Bei beiden wird den Teilnehmenden ein Musikstück vorgespielt, welches nach einer kurzen Übungszeit (10 Sekunden bei Melodien, 5 Sekunden bei Rhythmen, basierend auf Hasselhorn [Ha15]) auf dem im Projekt entwickelten ColorPiano (siehe Abb.2) nachgespielt werden soll. Das ColorPiano wurde in Anlehnung an die Colored Music Grid App von Hasselhorn [Ha15] gestaltet.

⁹ <http://www.musik.uni-bremen.de/de/forschung/musikpaedagogische-forschungsprojekte/komus-kompetenzmodell-im-fach-musik.html>

¹⁰ <http://alt.hfm-wuerzburg.de/kopra-m.html>

Bruder Jakob

Die folgende Melodie wird ein Mal vorgespielt, dabei kannst du sie dir anhören. Danach hast du 10 Sekunden Zeit zu üben. Am Ende der Zeit ertönt ein Gong. Hiernach wird die Melodie noch einmal gespielt und du sollst sie währenddessen auf dem Color Piano mitspielen.

● Aufgabe starten ⏱ 10 Sekunden

The image shows a musical score for 'Bruder Jakob' on a grand staff. The melody is presented in two staves. The notes are color-coded to correspond to the keys on a Color Piano. The first staff contains the melody, and the second staff contains the accompaniment. Below the staves, there are nine colored squares representing the keys: orange, grey, light green, red, yellow, green, cyan, brown, and purple. A 10-second timer icon is located at the top right of the score area.

Abb.2: Melodieitem zum Nachspielen auf dem ColorPiano

Die Musikstücke unterliegen dabei verschiedenen Schwierigkeitsgraden bzgl. z. B. der nachzuspielenden Stimme, des Aufbaus des Stückes oder der Abfolge der Noten. Um die potentielle Ungleichheit der Erfahrung von Schülerinnen und Schülern mit Instrumenten auszugleichen, wurde die Abspielmöglichkeit des ColorPiano derart gestaltet, dass die einzelnen Töne sowohl bei der Anzeige der Melodie als auch der vorhandenen Eingabefelder anhand von Farben dargestellt wurden. Hierbei wurde, soweit möglich, die Farbgebung barrierefrei gestaltet. Zur vereinfachten Angabe von Tonlängen wurden bei der Melodieanzeige zusätzlich Noten hinzugefügt. Um sicherzustellen, dass die Teilnehmenden den Überblick über den aktuellen Moment im Stück nicht verlieren, läuft während des Abspielens ein Slider mit, der die aktuelle Stelle kennzeichnet. Bei beiden Itemarten wird eingezählt, so dass sich die Teilnehmenden auf den Start des Stückes vorbereiten können. Nach Ablauf der Übungszeit wird ein Gong abgespielt, um das Ende dieser sowie den Beginn des Nachspielteiles zu kennzeichnen. Dieser musste sich sowohl von den Tönen aus den Stücken unterscheiden, als auch laut genug sein, damit er von den Teilnehmenden wahrgenommen werden kann. In Anbetracht des Kontextes Schule wurde hier versucht, einen Gong zu wählen, der einem Schulgong ähnelt und damit den Schülerinnen und Schülern bekannt ist. Bei den Melodieitems läuft während des Nachspielens die Melodie in einer anderen Klangfarbe mit, damit die Teilnehmenden sich zum einen an dieser orientieren können, zum anderen ihr eigenes Spielen aber nicht davon übertönt wird. Auch bei den Rhythmusitems wird während des Nachspielens eine Referenzmelodie abgespielt, um das Nachspielen zu erleichtern. Beispiele für beide Itemarten werden in Abb.2 sowie Abb.3 gegeben.

Rhythmus 41

Der folgende Rhythmus wird dir einmal vorgespielt, dabei kannst du ihn dir anhören. Danach hast du 5 Sekunden Zeit zu üben. Am Ende der Zeit ertönt ein Gong. Hiernach sollst du den Rhythmus zur Melodie auf dem Color Piano mitspielen.

⌚ 05 Sekunden

Abb.3: Rhythmusitem

4 Durchführung in den Schulen

4.1 Methodik

Zur Abwägung der für den Aufbau des Testsystems einzusetzenden Technik spielte die Erhebung der in den Schulen befindlichen technischen Infrastruktur eine große Rolle. Zur Bewertung dieser fand die Entwicklung eines kurzen Fragebogens mit nominal skalierten und offenen Fragen statt [Di14], welcher durch die Technikverantwortlichen der jeweiligen Schule beantwortet wurde. Die Auswertung erfolgte mit Hilfe einfacher Häufigkeiten für die nominal skalierten Fragen bzw. durch die Codierung der offenen Antwortblöcke, angelehnt an bekannte Inhaltsanalyseverfahren [GL04]. Auf Grundlage der gewonnenen Erkenntnisse folgte die Erstellung des Forschungsdesigns für die späteren Fallstudien in den Schulen [Yi06], [Ge07]. Die Leitung der Tests erfolgte durch die Projektmitarbeiterinnen und -mitarbeiter. Ein Mitglied der Testleitung führte hierbei eine teilnehmende Beobachtung durch [SW04]. Somit war es möglich, auf Fragen der Lernenden während des Tests einzugehen und bei Vorkommnissen, wie z. B. Absprachen von Lernenden, einzugreifen. Ein ggf. zweites Mitglied der Testleitung führte eine reine Beobachtung durch. Die daraus gewonnenen Erkenntnisse wurden nach den Tests in ein ausführliches Beobachtungsprotokoll inklusive der Skizzierung des Testaufbaus überführt. Mit der anwesenden Lehrkraft fand nach Beendigung des Tests eine kurze Retrospektive zu ihren Eindrücken bzgl. des Tests und des genutzten Systems statt. Diese Informationen wurden protokollarisch festgehalten. Anschließend fand die Besprechung und Beurteilung der gewonnenen Erkenntnisse aus jedem durchgeführten Test im Projektteam statt. Sofern notwendig konnten Anpassungen im Forschungsdesign für die nachfolgenden Fallstudien vorgenommen werden.

4.2 Durchführung des Tests

Die Durchführung des Tests war in allen teilnehmenden Schulen ähnlich. Unterschiede gab es zum einen bzgl. der Vorbereitung, zum anderen durch die Ergebnisse und Erfahrungen in den Schulen. Durch diese konnte der Ablauf bei den nachfolgenden Durchläufen anhand der zuvor gemachten Erfahrungen optimiert werden. Zu Beginn gab es eine kurze Einführung, in der die Testleitung sich und das Projekt vorstellte sowie kurz das Ziel des Tests erläuterte. Hierbei wurde explizit betont, dass die Tests nicht mit in die Benotung eingehen und außer der Lehrkraft niemand die Ergebnisse den Testpersonen zuordnen kann, die Tests also anonym durchgeführt werden. Um dies sicherzustellen, wurden den Schülerinnen und Schülern Handzettel ausgeteilt, auf denen Logindaten, bestehend aus einem Benutzernamen (Zahlen, Buchstaben und Sonderzeichen) und einem Passwort, standen. Um der Lehrkraft zu ermöglichen, später die Zuordnung zwischen Login und Lernenden vorzunehmen, wurden die Handzettel durch die Lernenden mit Namen versehen und am Ende der Stunde von der Lehrkraft eingesammelt. Die Zettel verblieben im Besitz der Lehrkraft.

Nach der allgemeinen Einführung wurde das Testsystem vorgestellt, indem die drei verschiedenen Aufgabentypen anhand von nicht im Test verwendeten Beispielaufgaben gezeigt und kurz erklärt wurden. Für die Einrichtung der Verbindung zum Testserver wurden in den Schulen verschiedene Varianten durchgeführt. In einigen Fällen standen die Tablets der Testleitung vor Beginn der Tests zur Verfügung. Die Verbindung konnte so ohne das Beisein der Testpersonen hergestellt werden, so dass nach der Einführung gleich mit dem Test begonnen werden konnte. War dies nicht der Fall, musste die Verbindung mithilfe der Lernenden hergestellt werden. Das erste Verfahren sah dabei vor, die Teilnehmenden nacheinander (einzeln bzw. in Zweiergruppen) zur Testleitung zu rufen. Diese richtete die Verbindung auf dem Tablet ein und teilte dazu den Handzettel aus. Da dieser Vorgang aber sehr zeitaufwendig war, wurde das Verfahren optimiert, indem die erforderlichen Schritte per Screenshots als Schritt-für-Schritt-Anleitung in die Präsentation eingefügt wurden. Da alle teilnehmenden Schulen iPads benutzen, musste nur eine Tabletvariante berücksichtigt werden. Dieses Vorgehen brachte eine enorme Verbesserung hinsichtlich der benötigten Zeit und stellte auch für die Schülerinnen und Schüler keine Schwierigkeit dar. Nach Herstellen der Verbindung mit dem Netzwerk wurden die Lernenden gebeten, die URL des Tests im Browser aufzurufen und sich dort mit den vorhandenen Accountdaten einzuloggen. Nicht auf allen Tablets funktionierten diese Schritte korrekt, so dass bei jedem Test ein paar Tablets ausgetauscht werden mussten. Leider konnte vom Projektteam nicht reproduziert werden, was diese Probleme verursacht. Da genug Geräte vorhanden waren, stellte dies aber kein Problem dar. Nach dem Einloggen wurden die Teilnehmenden aufgefordert, mit dem Übungstest zu starten. Hierbei starteten einige fälschlicherweise anstatt des Übungstests direkt den eigentlichen Test. Der Übungstest enthielt zwei Aufgaben aus dem Bereich Musikpraxis, um den Umgang mit dem ColorPiano ausprobieren zu können. Diese Übungsaufgaben wurden angelegt, da verschiedene Vortests zeigten, dass die Bedienung des ColorPianos beim ersten Mal nicht einfach ist und somit eine kurze Übungsphase wünschenswert wäre, um die Schülerinnen und Schüler nicht direkt zu

entmutigen und zu verschrecken. Somit wurde aus beiden Musikpraxisbereichen jeweils eine Aufgabe ausgewählt, mit denen die Bedienung ausprobiert werden konnte. Die Übungszeit war auf 5 bzw. 2 Minuten begrenzt. Nach Ablauf der Zeit wurde automatisch zur nächsten Übungsaufgabe weitergeleitet bzw. der Übungstest beendet. Die Schülerinnen und Schüler wurden aufgefordert, nach dem Übungstest mit dem eigentlichen Test zu starten. Bei einigen Schulen durften die Lernenden nach Beendigung des Tests den Raum verlassen, bei den meisten verblieben sie im Raum. Die zur Durchführung benötigte Zeit variierte stark. Einige Schülerinnen und Schüler schafften es in 20 Minuten, andere brauchten bis zu 45 Minuten. Auf dieses Phänomen und den dadurch steigenden Lautstärkepegel wurde im Laufe der Tests reagiert, indem Zusatzaufgaben hinzugefügt wurden, die die Lernenden nach Ablauf des eigentlichen Tests durchführen konnten. Somit konnte die Zeit, in der bei schnellen Teilnehmenden Langeweile aufkommen konnte sowie der Druck bei langsameren Teilnehmenden, den Test zu beenden, verringert werden. Sowohl vor Beginn des Tests als auch zwischendurch wurde durch die Testleitung sowie die Lehrkräfte darauf hingewiesen, dass die Tests einzeln und ohne Absprachen durchzuführen seien. Verständnisfragen sollten an die Testleitung gestellt werden. Soweit möglich wurde nach Beendigung des Tests ein Feedbackgespräch mit den Schülerinnen und Schülern geführt, in welchem sie gebeten wurden, ihre Erfahrungen und ihre Meinung zu dem Test zu äußern.

4.3 Ergebnisse aus den Testdurchführungen

Die Beobachtungen bei der Durchführung der Tests sowie die Aussagen der Teilnehmenden ähnelten sich über die Schulen hinweg. Im Großen und Ganzen machte der Test den Schülerinnen und Schülern Spaß. Sie empfanden ihn nicht als Zeitverschwendung und auch von der Länge her angemessen. In einer Schule, in der sehr viel mit Tablets gearbeitet wurde, wurde angemerkt, dass die Gestaltung des Tests etwas langweilig sei und aufregender gestaltet werden konnte. Die Durchführung der Musikpraxis-Items, besonders der Aufgaben, bei denen eine Melodie nachgespielt werden musste, wurde als schwierig empfunden. Die Stücke waren den Teilnehmenden teilweise zu schnell. Hier wünschten sich viele Schülerinnen und Schüler, dass sie die Aufgabe ein weiteres Mal durchführen könnten. Teilweise versuchten sie dies, indem sie die Browser Zurück-Funktion nutzten. Da dieses Verhalten im Testsystem nicht vorgesehen ist, hatte dies zur Folge, dass die Jugendlichen wieder auf die Startseite geleitet wurden und, nach Hilfe durch die Testleitung, den Test neu starten bzw. fortsetzen mussten. Zudem wurde angemerkt, dass die Übungszeit für die einzelnen Items zu kurz sei und es nicht möglich wäre, in dieser Zeit das ganze Stück einmal durchzuspielen. Hier hätten sich viele eine längere Übungszeit gewünscht. Hingegen wurde die Übungszeit am Anfang während des Übungstests als zu lang empfunden. Dies sorgte dafür, dass einige Schülerinnen und Schüler unruhig wurden oder das Tablet zur Seite legten und darauf warteten, dass die Zeit abliefe. Bei den Rhythmusitems wurde angemerkt, dass die Begleitmusik zum einen zu laut, zum anderen zu schnell und zudem irritierend war, da sie nicht beim ersten Abspielen, sondern erst in der Nachspielphase abgespielt wird. Hierauf wurde im Laufe der Tests reagiert, indem die Lautstärke angepasst wurde und

somit die gespielten Töne deutlicher vor der Hintergrundmusik wahrgenommen werden können. Die Bedienung wurde generell als einfach empfunden. Ein Schüler aber merkte an, dass die Kästchen zum Spielen zu eng beieinander sind. Andere Schülerinnen und Schüler empfanden die Farben des ColorPianos verwirrend, auch insofern, dass in den unterschiedlichen Stücken unterschiedliche Farben für die Noten verwendet wurden. Zudem würden sie sich Noten auf den Kästchen wünschen. Hier gab es von einigen Jugendlichen die Rückmeldung, dass sie sich generell an den Noten orientiert haben. Dies sei für sie einfacher, da sie auch Instrumente spielen und es von daher so gewohnt sind. Bei den Textaufgaben gab es geteilte Meinungen. Einigen Teilnehmenden machten diese mehr Spaß als die Praxisaufgaben. Die meisten empfanden die Aufgaben als „schaffbar“ und nicht zu schwer. Andere dachten aber auch, dass sie sie nicht schaffen würden. Die Aufgabe zum freien Assoziieren wurde als besonders schwer empfunden. Dies konnte sowohl am Verhalten der Teilnehmenden wahrgenommen werden, da diese unruhiger wurden und z. B. bei ihren Nachbarinnen und Nachbarn schauten. Zum anderen formulierten es auch einige in der Nachbesprechung. Interessant war auch, dass es als schade empfunden wurde, dass es keine direkten Rückmeldungen zu den Aufgaben gab.

5 Fazit und Ausblick

Die Durchführung der Tests in den Schulen zeigt auf, dass der im Projekt verfolgte Ansatz zukunftsfähig ist. Sowohl die Schülerinnen und Schüler als auch die Lehrkräfte gaben größtenteils positives Feedback und standen dem Einsatz eines solchen Testsystems offen gegenüber. Ein Lehrer äußerte sogar den Wunsch, das System fortwährend in seinem Unterricht einzusetzen, um somit die Musikkompetenzen seiner Schülerinnen und Schüler auch über einen längeren Zeitraum hinweg messen und vergleichen zu können. Gleichzeitig wurde auch Verbesserungspotential deutlich. Zum einen stellte die Bedienung des ColorPianos die Testpersonen vor Schwierigkeiten. Dieser Effekt könnte durch einen fortwährenden Einsatz vermutlich gemindert werden. Ebenso hilfreich wäre hier z. B. zu ermöglichen, die Geschwindigkeit der Musikstücke zu beeinflussen oder eigene Testaufgaben zu erstellen. Zudem wurde auch der Wunsch nach Rückmeldungen zu den bearbeiteten Aufgaben deutlich. Diesem Bedürfnis wird im Projekt PopsyMus mit der Entwicklung eines elektronischen Rückmeldesystems begegnet, welches die Lehrkräfte bei der Gestaltung ihres Unterrichts unterstützen soll, indem es ermöglicht, die gemessenen Musikkompetenzen anhand von verschiedenen Visualisierungen interaktiv zu betrachten.

Literaturverzeichnis

- [AMP04] Abras, Chadia; Maloney-Krichmar, Diane; Preece, Jenny: User-centered design. In (Bainbridge, William, Hrsg.): Encyclopedia of Human-Computer Interaction 37/4, Thousand Oaks: Sage Publications, London, S. 445-456, 2004.

- [As11] Astafan, Christine: SMARTMUSIC: Using Technology to Assess Rhythmic Ability within Instrumental Music in the Elementary School Classroom. Online Submission, 2011.
- [BTP99] Balizet, Sha; Treder, Dave; Parshall, Cynthia G.: The Development of an Audio Computer-Based Classroom Test of ESL Listening Skills. In: Annual Meeting of the American Educational Research Association. Montreal, 1999.
- [Di14] Diekmann, Andreas: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Rowohlt-Taschenbuch-Verl., Reinbek bei Hamburg, 2014.
- [Do13] Doepfert, Julia: Erfassung musikpraktischer Kompetenzen am Instrument - ein experimenteller Vergleich von Keyboard oder Tablet-App als alternative Eingabebereiche. Schriftliche Hausarbeit zur 1. Staatsprüfung für das Lehramt Gymnasien. Hochschule für Musik Würzburg, 2013.
- [FMB16] Finken, Julia; Marx, Franziska; Breiter, Andreas: Computerbasiertes Testen zur Messung von Musikkompetenzen. In (Lucke, Ulrike; Schwill, Andreas; Zender, Raphael, Hrsg.): Proceedings of DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V., Gesellschaft für Informatik, Bonn, S. 327–329, 2016.
- [Ge07] Gerring, J.: Case Study Research: Principles and Practices. Cambridge Univ. Press, Cambridge, 2007.
- [GH06] Geisser, Michael; Hildenbrand, Tobias: Eine Methode zur kollaborativen Anforderungserhebung und entscheidungsunterstützenden Anforderungsanalyse. Universität Mannheim, Fakultät für Informationssysteme, Mannheim, 2006.
- [GI00] Glas, Gees A. W.: Item Calibration and Parameter Drift. In (van der Linden, Wim J.; Glas, Gees A. W., Hrsg.): Computerized Adaptive Testing: Theory and Practice, Springer Netherlands, S. 183-199, 2000.
- [GL04] Glaeser, Jochen; Laudel, Grit: Experteninterviews und qualitative Inhaltsanalyse. VS Verlag für Sozialwissenschaften, Wiesbaden, 2004.
- [Ha15] Hasselhorn, Johannes: Messbarkeit musikpraktischer Kompetenzen von Schülerinnen und Schülern: Entwicklung und empirische Validierung eines Kompetenzmodells, Jgg. 2 in Perspektiven musikpädagogischer Forschung. Waxmann, Münster, 2015.
- [HL14] Hasselhorn, Johannes; Lehmann, Andreas: Entwicklung eines empirisch überprüfbaren Modells musikpraktischer Kompetenz (KOPRA-M). In (Clausen, Bernd, Hrsg.): Teilhabe und Gerechtigkeit, Waxmann, Münster, S. 77-94, 2014.
- [JH07] Jurecka, Astrid; Hartig, Johannes: Computer- und netzwerkbasiertes Assessment. In (Hartig, Johannes; Klieme, Eckhard, Hrsg.): Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Eine Expertise im Auftrag des Bundesministeriums für Bildung und Forschung, Bonn, S. 37-48, 2007.
- [JKL10] Jordan, Anne-Katrin; Knigge, Jens; Lehmann-Wermser, Andreas: Projekt KoMus: Entwicklung von Kompetenzmodellen in einem ästhetischen Fach. In (Gehrmann, Axel; Hericks, Uwe; Luders, Manfred, Hrsg.): Bildungsstandards und Kompetenzmodelle - Beiträge zu einer aktuellen Diskussion über Schule, Lehrerbildung und Unterricht, Klinkhardt, Bad Heilbrunn, S. 209–222, 2010.
- [Jo12] Jordan, Anne-Katrin; et al.: Entwicklung und Validierung eines Kompetenzmodells im

- Fach Musik - Wahrnehmen und Kontextualisieren von Musik. *Zeitschrift und Pädagogik* 58/4, S. 500–521, 2012.
- [JW07] Jude, Nina; Wirth, Joachim: Neue Chancen bei der technologiebasierten Erfassung von Kompetenzen. In (Hartig, Johannes, Hrsg.): *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*. Jgg. 20, BMBF, Berlin, S. 49–56, 2007.
- [Ma09] Martin, Ron: Utilising the potential of computer delivered surveys in assessing scientific literacy. In (Björnsson, Julius; Scheuermann, Friedrich, Hrsg.): *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-scale Testing*. Office for Official Publications of the European Communities, Luxembourg, 2009.
- [Ni93] Nielsen, Jakob: *Usability Engineering*. Morgan Kaufmann, San Francisco, 1993.
- [Ni08] Niessen, Anne; et al.: Entwurf eines Kompetenzmodells “Musik wahrnehmen und kontextualisieren”. In (Hartig, Johannes; Klieme, Eckhard, Hrsg.): *Zeitschrift für Kritische Musikpädagogik*, S. 3-33. 2008.
- [SK08] Staphorsius, Gerrit; Krom, Ronald: Das Schüler-Monitoring-System in den Niederlanden. In (Berkemeyer, Nils; Bos, Wilfried; Manitus, Veronika; Müthing, Kathrin, Hrsg.): *Unterrichtsentwicklung in Netzwerken. Konzeptionen, Befunde, Perspektiven*, Waxmann, Münster, S. 151-182, 2008.
- [St12] Stauke, Emese: *Konzeption und Anforderungsanalyse eines elektronischen Rückmeldesystems für schulische Leistungsdaten*. Dissertation, Universität Bremen, Bremen, 2012.
- [SW04] Schonhagen, P.; Wittmann, F.: *Teilnehmende Beobachtung: Datenerhebung “hautnah” am Geschehen*. Institut des sciences économiques et sociales. FSES, 2004.
- [VWB97] Vispoel, Walter P.; Wang, Tianyou; Bleiler, Timothy: Computerized Adaptive and Fixed-Item Testing of Music Listening Skill: A Comparison of Efficiency, Precision, and Concurrent Validity. *Journal of Educational Measurement* 34/1, S. 43-63, 1997.
- [We01] Weinert, Franz E.: *Leistungsmessungen in Schulen*. Weinheim u.a., 2001.
- [Ya08] Yap, Ching Ching: A statewide web-based music assessment: a collaborative effort. In (Brophy, Timothy S., Hrsg.): *Assessment in Music Education: Integrating Curriculum, Theory, and Practice*. GIA Publications, Chicago, 2008.
- [Yi06] Yin, R. K.: *Case Study Research: Design and Methods*. Sage, Thousand Oaks, CA, 2006.
- [Ze15] Zentrum für internationale Vergleichsstudien: *PISA 2015. Bericht*, OECD/TUM School of Education, München, 2015.

Evaluation einer Statistiklehrveranstaltung mit dem JACK R-Modul

Benjamin Otto¹, Till Massing², Nils Schwinning¹, Natalie Reckmann², Alexander Blasberg², Sandy Schumann², Christoph Hanck² und Michael Goedicke¹

Abstract: Neben klassischen Multiple-Choice oder Lückentextaufgaben bieten moderne elektronische Lernsysteme die Möglichkeit, Programmieraufgaben automatisiert zu bewerten. Das E-Assessment System JACK wurde um einen Aufgabentyp für Programmieraufgaben mit der statistischen Programmiersprache R ergänzt, da diese innerhalb der statistischen Gemeinschaft weit verbreitet ist. Sie wird daher auch Studierenden der Wirtschaftswissenschaften an der Universität Duisburg-Essen bereits in der Studieneingangsphase vermittelt. Dieser Artikel schildert die Nutzung des neuen Aufgabentyps in einem Erstsemester-Statistikurs und zeigt, dass ein positiver Effekt von Programmieraufgaben auf klassische Klausuraufgaben festzustellen ist.

Keywords: E-Assessment, R, Statistik, Lehrveranstaltung, Evaluation, Autograder

1 Einleitung

Der Einsatz von elektronischen Übungssystemen in der Studieneingangsphase ist ein weit verbreitetes Konzept, um wachsenden Studierendenzahlen angemessen begegnen zu können. Der in vielen Studienfächern übliche, klassische Übungsbetrieb wird dabei durch Online-Assessments ersetzt, um Studienanfängern die Möglichkeit zu geben Lerninhalte zu üben. Neben der schnellen Rückmeldung in Form von automatischen Bewertungen spielt auch die hohe Flexibilität bei dem Einsatz von E-Assessment Systemen eine Rolle, denn Studierende sind in der Lage beliebig oft und zu jeder Tageszeit mit ihnen zu arbeiten. Konzepte reichen von klassischen Aufgabentypen, wie Multiple-Choice-Aufgaben³ oder Lückentextaufgaben⁴ bis hin zu Programmieraufgaben, z.B. mit Java [St16].

In den Vorlesungen der Statistik an der Universität Duisburg-Essen wird das selbst entwickelte E-Assessment-System JACK zur Unterstützung des Vorlesungsbetriebs eingesetzt. Das Konzept beinhaltet bisher wöchentliche formative und summative Assessments in Form von parametrisierten Lückentextaufgaben [Sc17]. Analysen haben gezeigt, dass die Note der Abschlussklausur mit dem Lernerfolg bei diesen Assessments

¹ Universität Duisburg-Essen, Paluno - The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, vorname.nachname@paluno.uni-due.de

² Universität Duisburg-Essen, Wirtschaftswissenschaften, Universitätsstr. 12, 45117 Essen, vorname.nachname@vwl.uni-due.de

³ Z.B. Moodle: <https://moodle.de/>

⁴ Z.B. Moodle, Stack: <http://stack.bham.ac.uk/>, ActiveMath: <http://www.activemath.org/>

positiv korreliert [Ma17]. Jedoch stoßen die Lückentextaufgaben bei komplexeren Problemstellungen an ihre Grenzen. So spielt die Programmiersprache R [RC17] in der Statistikausbildung eine wichtige Rolle und dient als Hilfswerkzeug für viele Berechnungen. Studierende lernen daher bereits in Einführungsveranstaltungen den Umgang mit der Sprache, da ihnen dadurch die Anwendung der erlernten Methoden erleichtert wird. Es wäre somit wünschenswert, auch die R-Programmierung durch E-Assessments unterstützen zu können.

Dieser Artikel beschreibt einen neuen Aufgabentyp, welcher es ermöglicht, Aufgaben zu stellen, die mit R zu lösen sind. Der Aufgabentyp wurde dem E-Assessment-System JACK hinzugefügt und im Wintersemester 2016/17 in das bestehende Konzept von semesterbegleitenden formativen und summativen E-Assessments in einer Vorlesung zur Deskriptiven Statistik integriert. Lernende müssen dabei zu einer vorgegebenen Aufgabenstellung R-Code schreiben und können diesen anschließend hochladen. Das System prüft diese Lösung mittels dynamischer und statischer Tests und gibt den Lernenden ein detailliertes Feedback. Es wurden insgesamt 72 Aufgaben für die Vorlesung erstellt und eingesetzt. Wir evaluieren die Nutzung dieser Aufgaben durch die Studierenden und beschreiben ihren Lernerfolg hinsichtlich der Abschlussklausur.

Der Rest dieses Artikels gliedert sich wie folgt: Zunächst werden in Kapitel 2 verwandte Arbeiten beschrieben. Kapitel 3 erläutert den neuen Aufgabentyp und Kapitel 4 das Konzept der Vorlesung sowie den Einsatz der verschiedenen Assessmentformen. In Kapitel 5 wird das Lernverhalten der Studierenden während des Semesters analysiert und mit den Ergebnissen der Abschlussklausur verglichen. Zum Schluss wird in Kapitel 6 ein Fazit gezogen und ein Ausblick auf weitere Forschung gegeben.

2 Verwandte Arbeiten

Der Einsatz von elektronischen Übungssystemen in universitären Lehrveranstaltungen kann sowohl das Themenverständnis als auch die Leistung in Prüfungen der Studierenden erheblich verbessern [CM02], [BB14], [Me13]. Für den Bereich der Statistik hat [So11] dies in einer Meta-Analyse ausführlich untersucht und kommt zu dem Ergebnis, dass die simultane Verwendung von Präsenzveranstaltung und Online-Assessments einen positiven Effekt auf den Lernerfolg von Studierenden hat. Insbesondere bei datenintensiven Disziplinen zeigt sich jedoch, dass eine alleinige Vermittlung von traditionellen Lehrbuchinhalten heute nicht mehr ausreicht und die Datenanalyse sowie -aufbereitung mit entsprechenden Software-Programmen eine wichtige zu beherrschende Fähigkeit darstellt [TL12], [Ch07].

In ersten Studien konnte gezeigt werden, dass die Verwendung von statistischen Softwareprogrammen zu Analyse- und Visualisierungszwecken das Verständnis für komplexe Zusammenhänge noch weiter verbessert und bessere Prüfungsergebnisse erzielt werden. So erforschte [Ba05] in einer einführenden Statistiklehrveranstaltung, ob Studierende mit Hilfe des Software Programms SPSS ein besseres statistisches

Verständnis erhalten. In der abschließenden Prüfung erreichten Studierende, die während des Semesters wöchentlich in einem Computerlabor Zugriff auf SPSS hatten, eine höhere Punktzahl als jene, die nur die klassische Vorlesung besuchten. In diesem Kontext erscheint es sinnvoll, das Lernen der theoretischen Konzepte auf der einen Seite mit der praktischen Anwendung dieser auf der anderen Seite zu verknüpfen. Entsprechende Möglichkeiten, wie diese Verknüpfung in einer E-Assessment-Umgebung realisiert werden kann, werden von [Mc15] ausführlich vorgestellt. Neben der konkreten Ausgestaltung des Übungssystems stellt sich hierbei die Frage nach der Wahl des zu verwendenden Programms. Aufgrund seiner Popularität, Flexibilität sowie Erweiterbarkeit bietet sich hierfür R an [Mu15]. Eine entsprechende Lehrplattform existiert in der Form von datacamp [Da17] bereits, allerdings liegt der Hauptfokus dort eher auf dem Aspekt des Selbstlernens und weniger auf dem summativen Assessment. Eine entsprechende, auf Hochschullehre zugeschnittene, E-Assessment-Plattform mit integrierter R-Anbindung existiert nach bestem Wissen der Autoren noch nicht.

3 Der JACK-R-Aufgabentyp

Im Rahmen des ProViel-Projekts zur Verbesserung der Lehramtsausbildung wurde ein neuer Aufgabentyp entwickelt, der Aufgaben für die statistische Programmiersprache R im E-Assessment-System JACK ermöglicht. In diesem Aufgabentyp sehen die Studierenden zunächst eine Aufgabenstellung, welche in R ausgerechnete (z.B. randomisierte) Werte enthalten kann. Hierbei steht den Studierenden ein Texteingabefeld zur Verfügung, welches sowohl der Eingabe und dem Einreichen der Lösung dient, als auch als Texteingabefeld für die Live-R-Konsole. Diese ermöglicht es (ohne Installation zusätzlicher Software) den Studierenden R-Programme im Browser zu schreiben und auch live auszuwerten. Dazu steht ein Editor mit Syntax-Highlighting zur Verfügung. Durch Klicken des Buttons „Auswerten“ wird der gesamte angezeigte Code an einen R-Server gesendet, der die Programmausführung übernimmt und die Rückgaben in das Ausgabefeld einträgt. So sind Studierende in der Lage, ihre Lösungen direkt auf dem Server zu testen und sich die Rückgaben anzusehen. Die Studierenden bekommen dadurch sofort einen Eindruck davon, ob die Syntax ihres Programms korrekt ist, da alle in R auftretenden Fehler an die Benutzeroberfläche durchgereicht werden. Aber es kann auch unverzüglich gesehen werden, ob die Programmausgabe einerseits das erwartete Format zurückgibt oder andererseits im erwarteten Zahlenbereich liegt. Durch die schnelle und einfache Berechnung von R-Befehlen im Browser sollen die Studierenden zum Ausprobieren angeregt werden.

In Abb. 3.1 ist eine typische Aufgabe dargestellt. In dieser Aufgabe sollen die Quintile eines Datensatzes mit R dargestellt werden. Hier hat ein Studierender zwar die richtige R-Funktion (`quantile`) gefunden, aber dieser wurden noch nicht die richtigen Parameter für die Aufgabenstellung (`probs=seq(0,1,0.2)`, `type=1`) übergeben. Da der Studierende den Button „Auswerten“ geklickt hat, ist im unteren Textfeld die Ausgabe von R zu sehen. Eine Besonderheit ist hier, dass dem Studierenden schon

vorausgefüllter Code (der sog. „Initial-Code“ `urliste <- c(11, ..., 15)`) in dem Texteingabefeld präsentiert wurde. Dieser kann optional vom Lehrenden angegeben werden und auch dynamisch berechnete Werte enthalten.

Aufgabe "(Aufgabe05) Quantile"

Lassen Sie sich mit Hilfe von R die Quintile des gegebenen Datensatzes *urliste* ausgeben. (Wichtiger Hinweis: Geben Sie den Parameter `type=1` an! Sonst werden die Quantile anders berechnet (die Default-Einstellung ist `type=7`))

```
1 urliste <- c(11,13,15,16,12,18,14,15,17,14,12,16,13,15,17,16,15,14,13,15)
2 quantile(urliste)
```

Hinweis **Auswerten** **Abschicken**

```
1 [1] 11 13 15 16 12 18 14 15 17 14 12 16 13 15 17 16 15 14 13 15
2
3 0% 25% 50% 75% 100%
4 11 13 15 16 18
```

Abb. 3.1: Aufbau einer Jack-R-Aufgabe

Automatische Bewertung der Benutzereinreichungen und Feedback

Um große Anzahl an Benutzereinreichungen sowohl automatisch zu bepunkten, als auch dem Lernenden ein hilfreiches Feedback anbieten zu können, bedienen wir uns zweier Testtechniken aus der Softwaretechnik, dem statischen und dem dynamischen Testen von Programmen [KJH16]. Beim statischen Test (White-Box-Test) können Aussagen über die interne Struktur des Codes gemacht werden. Im dynamischen Test (Black-Box-Test) können im Gegensatz dazu Aussagen über die Funktionalität des Codes gemacht werden.

3.1 Statischer R-Checker

Zunächst wurde für die Entwicklung des statischen R-Checkers eine R-Grammatik entworfen, auf deren Basis R-Programme in eine Graphenstruktur geparkt werden können. Auf den so generierten Graphen können nun mit GReQL [Un10] von Lehrenden beliebige Graphabfragen formuliert werden.

Auf diese Art kann vom Lehrenden beispielsweise überprüft werden, ob Studierende die in der Lösung gewünschten Funktionen mit korrekten Argumenten genutzt haben, ob Benamungskonventionen eingehalten wurden oder ob Schleifenkonstrukte zur Lösung genutzt wurden.

3.2 Dynamischer R-Checker

Für den dynamischen Checker kann der Lehrende einen oder mehrere Testfälle anlegen. Jeder Testfall vergleicht die Ausgabe der Studierendeneinreichung mit dem erwarteten Ergebnis. Zum Bestehen oder Fehlschlagen jedes Testfalls kann ein Feedback von Lehrenden angegeben werden.

R wird in der betrachteten Veranstaltung eher wie eine Skriptsprache benutzt, so dass das klassische Konzept des Unittestings auf von Usern geschriebenen Unterfunktionen nur bedingt greifen kann. Diese basieren darauf, dass Userfunktionen mit verschiedenen Argumenten aufgerufen werden und die Rückgabewerte auf vorher bestimmte Werte geprüft werden. Man kann sich so nach einer bestimmten Anzahl von Tests sicher sein, dass die Funktion richtig funktioniert. Deshalb generiert der R-Checker zunächst eine R-Unterfunktion aus der Einreichung. Diese kann dann wie im klassischen Unittesting mit den in den Testfällen definierten Übergabeparametern und den entsprechenden erwarteten Ergebnissen verglichen werden. Zum Vergleich der Userergebnisse mit den erwarteten Ergebnissen verwendet der dynamische Checker das Testthat-R-Modul [Wi16].

Ferner kann der Lehrende nach der generierten Testfunktion eigenen Code einfügen (postcode), so dass es möglich ist, dort noch vom Lehrenden benötigte Nachverarbeitungen durchzuführen. Dies ist hilfreich, wenn mehrere mögliche Antworten als richtig erkannt werden sollen oder die Antworten verschiedene Formate haben dürfen (z.B. eine Matrix ohne und eine mit Spaltenbezeichnungen).

In Abb. 3.2 ist Feedback zu einer Einreichung eines Studierenden zu sehen, bei dem sowohl Checker-Regeln des statischen, als auch die des dynamischen Checkers zu Fehlern geführt haben.

Ergebnisübersicht

Gesamtergebnis: 0
Mindestergebnis für eine korrekte Lösung ist 50

Dynamic R Checker (1) result

Einige Testfälle nicht bestanden, siehe unten!

DETAILLIERTE KOMMENTARE

(-) 1
Leider nicht richtig! Achten Sie darauf, dass Sie die Quantile als Vektor übergeben und den Datensatz in 5 Intervalle aufgeteilt haben.

Static R Checker (1) result

(-) Fehlerhafte Codestruktur
Leider nicht richtig! Achten Sie darauf, dass sie die *quantile*-Funktion benutzen und *type = 1* setzen.

Abb. 3.2: Automatisch generiertes Feedback nach der Einreichung aus Abb. 3.1

4 Didaktisches Konzept

Im Wintersemester 2016/17 wurde JACK im Rahmen der Lehrveranstaltung „Deskriptive Statistik“ an der Universität Duisburg-Essen am Lehrstuhl Ökonometrie eingesetzt. Die Hörschaft der Veranstaltung setzte sich zusammen aus zu Beginn insgesamt etwa 1000 Studierenden im 1. bis 3. Semester der Bachelorstudiengänge VWL, BWL, Wirtschaftsinformatik und Lehramt Wirtschaftswissenschaften. Von diesen 1000 Studierenden schrieben jedoch nur 457 eine der Klausuren. Lag der Fokus in früheren Semestern alleine auf der Vermittlung und der nicht-computerbasierten Anwendung von Methoden zur statistischen Datenanalyse, so hatte die Veranstaltung im Wintersemester 2016/2017 zusätzlich zum Ziel, die Anwendung der in Vorlesung und Übung erlernten Methoden auch computerbasiert, explizit mit der Statistiksoftware R, zu vermitteln. Zur Umsetzung dieses Ziels wurde das bestehende didaktische Konzept durch die Bereitstellung von statistischen Programmieraufgaben auf der Lernplattform JACK erweitert.

Konzeptionell setzte sich die Veranstaltung „Deskriptive Statistik“ zusammen aus einer wöchentlich stattfindenden Vorlesung, in der die statistischen Verfahren gelehrt werden, sowie einer ebenfalls wöchentlich stattfindenden Übung, in der die erlernten Methoden in praktischen Anwendungen vertieft werden. Beide Veranstaltungen fanden in Form von Frontalunterricht global für die gesamte Hörschaft statt. Aufgrund der Masse an Studierenden konnte hier nicht, oder nur sehr schwer, auf individuelle Fragen und ein unterschiedliches Lerntempo eingegangen werden. Um dieses Problem abzumildern und das aktive, selbständige Lernen der Studierenden zu fördern, wurden, wie bereits in vergangenen Semestern, Aufgaben zum selbständigen Bearbeiten auf der Lernplattform JACK bereitgestellt. Dieses formative Assessment wurde im aktuellen Semester durch Programmieraufgaben in der Programmiersprache R erweitert. Insgesamt wurden 172 Übungsaufgaben freigeschaltet, davon 100 Übungsaufgaben zu klassischem Vorlesungsstoff ohne Programmierinhalte. Die restlichen 72 Aufgaben wurden als R-Aufgaben konzipiert. Durch die Bereitstellung von Feedback zu häufig begangenen Fehlern beim Lösen einer Aufgabe sowie optional wählbarer Hinweise wird der individuelle Lernerfolg der Studierenden unterstützt. Bei eventuell auftretenden Fragen und Problemen bei der Bearbeitung der JACK-Aufgaben stand ein vom Lehrstuhl betreutes Onlinetutorium zur Verfügung. Es konnte zudem ein wöchentlich stattfindendes freiwilliges Tutorium besucht werden.

Zur Unterstützung einer kontinuierlichen Arbeitsweise der Studierenden wurden verteilt über die gesamte Vorlesungszeit insgesamt 5 Testate abgehalten, für die ebenfalls JACK genutzt wurde. Diese Testate hatten eine Dauer von 40 Minuten und fanden zu vorgegebenen Zeiten statt. Zur Teilnahme war lediglich ein internetfähiger PC erforderlich, es bestand aber keine Anwesenheitspflicht in Räumen der Universität. Dieses summative Assessment bot den Studierenden die Möglichkeit, ihren persönlichen Wissensstand kontinuierlich über die Vorlesungszeit hinweg zu überprüfen. Wie auch im Wintersemester 13/14 bildeten Bonuspunkte den Anreiz zur Ablegung dieser Testate, maximal 2,4 pro Testat, die im Falle einer am Semesterende bestandenen Klausur,

angerechnet wurden. Dies bot den Studierenden die Möglichkeit, maximal 12 Bonuspunkte für die Klausur zu sammeln, in der Klausur wurden jedoch maximal 10 Punkte angerechnet. Ein zusätzliches 6. Testat am Ende der Vorlesungszeit beinhaltete, im Gegensatz zu den vorherigen Testaten, R-Aufgaben, die ebenfalls zu einer festgelegten Zeit auf der Lernplattform JACK bereitgestellt wurden. Die Anrechnung von maximal 2 Bonuspunkten auf die am Semesterende abzulegende Klausur auch im Falle des Nichtbestehens bildete einen zusätzlichen Anreiz.

Die abschließende Klausur hatte eine Dauer von 70 Minuten und wurde im ersten Termin als PC-gestützte Klausur auf der Plattform JACK abgelegt. Im zweiten Termin wird eine gewöhnliche handschriftliche Klausur angeboten. In einem zusätzlichen dritten Termin wird den Studierenden die Möglichkeit geboten, erneut an einer PC-gestützten Klausur auf der Lernplattform JACK teilzunehmen. Die Klausuren im zweiten und dritten Termin wurden zum jetzigen Zeitpunkt noch nicht abgelegt. Im Gegensatz zu den abschließenden Klausuren in vorherigen Semestern bieten die Klausuren auf der Lernplattform JACK die Möglichkeit, zusätzlich R-Aufgaben abzufragen. So bestand die erste JACK-gestützte Klausur aus insgesamt 60 erreichbaren Punkten, davon konnten 52 Punkte durch klassische und 8 Punkte durch R-Aufgaben erzielt werden. Aufgrund der Unsicherheit bzgl. des Lernerfolgs der Studierenden bzgl. des in diesem Semester neu eingeführten Konzepts, R in das Lehrangebot aufzunehmen, wurde davon abgesehen, einen größeren Teil der Klausur durch R-Aufgaben abzudecken.

5 Evaluation

In der in Kapitel 4 beschriebenen Lehrveranstaltung wurden die Nutzerdaten auf JACK protokolliert und ausgewertet. An der ersten Abschlussklausur nahmen 329 Studierende teil, deren Lernerfolg hier analysiert werden soll. Wie in Kapitel 4 erläutert, bestand die Abschlussklausur zum Vortermin aus insgesamt 60 Punkten, davon konnten 8 Punkte durch eine R-Aufgabe erreicht werden. Wir untersuchen hier die Fragestellung, ob Studierende mit hohem Arbeitseinsatz in R-Übungsaufgaben davon auch in den Standardklausuraufgaben ohne R-Inhalte profitieren. Zur Kontrolle, ob der Effekt alleine aus den R-Aufgaben resultiert, betrachten wir zusätzlich auch die Einreichungen in den klassischen JACK-Aufgaben.

Insgesamt haben 457 Studierende an mindestens einer Klausur teilgenommen. Wir konzentrieren uns in nachfolgender Auswertung auf den ersten Termin, um die unterschiedlichen Modalitäten (elektronisch o. schriftlich, 3 Termine in WS16/17 u. 2 schriftliche Termine in WS14/15) nicht zu stark zu vermischen. Die Ergebnisse sind ohnehin zahlenmäßig sehr ähnlich und sind auf Anfrage verfügbar.

Von den Klausurteilnehmern liegen u.a. die folgenden Daten vor: Die Klausurpunkte der klassischen Aufgaben, Klausurpunkte der R-Aufgabe, die Note (setzt sich aus der Summe der Klausurpunkte und der Bonuspunkte zusammen), die Anzahl der Einreichungen zu JACK Aufgaben ohne R und die Anzahl der Einreichungen zu JACK-

R-Aufgaben. Zusätzlich haben wir jeweils die Anzahl an korrekten Einreichungen auf JACK erhoben. Diese korrelieren mit der jeweiligen Anzahl an Einreichungen zu jeweils circa 97%. Wir betrachten im Folgenden die Variablen zu allen Einreichungen, da dies als ein aussagekräftigeres Maß für studentischen Lerneinsatz erscheint als lediglich korrekte Einreichungen. Dieselbe Analyse unter Verwendung der korrekten Einreichungen führt aufgrund der hohen Korrelation zu sehr ähnlichen Ergebnissen.

Tab. 5.1 zeigt den um Zwischennoten bereinigten Notenspiegel des Jahrgangs Wintersemester 2016/17 im Vergleich mit den Noten des Jahrgangs 2013/14, beide jeweils zum Vorterrmin.

Noten (bereinigt)	Wintersemester 13/14	Wintersemester 16/17
1	22 (5,54%)	39 (11,85%)
2	60 (15,11%)	46 (13,98%)
3	124 (31,23%)	80 (24,32%)
4	27 (6,8%)	15 (4,56%)
5	164 (41,3%)	149 (45,23%)
Summe	397	329
Schnitt (unbereinigt)	3,64	3,59

Tab. 5.1: Notenvergleich für zwei Jahrgänge. Die Noten wurden zur besseren Übersicht um Zwischennoten wie 1,3 bereinigt. Der Notenschnitt bezieht sich auf die unbereinigten Noten.

Im Schnitt haben sich die Noten im Jahrgang mit R-Nutzung leicht verbessert, insbesondere, da sich die Anzahl an sehr guten Studierenden fast verdoppelt hat. Dafür hat sich der Anteil an nicht bestandenen Klausuren von 41,3% auf 45,23% leicht erhöht. Anscheinend wurden durch die Nutzung von R im aktuellen Semester zwar die guten Studierenden gefördert, dafür konnten schwache Studierende eher nicht davon profitieren. Diese Aussagen sind insgesamt jedoch mit Vorsicht zu verstehen, da es sich nur um sehr geringfügige Änderungen handelt.

Die auf JACK bereitgestellten Aufgaben wurden von den Studierenden 97.244 mal bearbeitet. Dies teilt sich auf in 67.015 Einreichungen zu Aufgaben ohne R und 30.229 Einreichungen zu R-Aufgaben. Der Anteil an bearbeiteten R-Aufgaben bzgl. der Anzahl aller Einreichungen betrug also circa 31%.

Abb. 5.1 zeigt Boxplots zur Anzahl an Klausurpunkten ohne die R-Aufgabe, Anzahl an bearbeiteten klassischen Aufgaben auf JACK und der Anzahl an bearbeiteten R-Aufgaben auf JACK je Student. Vergleicht man den mittleren mit dem rechten Boxplot, erkennt man, dass ein größerer Anteil der Studierenden die R-Übungsaufgaben weitestgehend ignoriert hat. Die Ausreißer sind dadurch zu erklären, dass ein Studierender die gleichen Aufgaben sehr oft wiedereingereicht hat.

Wir untersuchen nun den Effekt der R-Aufgaben auf den Lernerfolg und insbesondere auf den Klausurerfolg. Interessant ist hier vor allem, welchen Einfluss die R-Übungsaufgaben auf Klausuraufgaben ohne R-Inhalte haben. Im Falle eines positiven

Effektes würden die R-Aufgaben also auch das allgemeine statistische Verständnis fördern. Hierzu regressieren wir die Variable „Klausurpunkte ohne R“ (minimal 0, maximal 52 Punkte) auf die Variable „Anzahl an Einreichungen in R-Aufgaben“ und kontrollieren für die Variable „Anzahl an Einreichungen in klassischen Aufgaben“ mittels der Standard Kleinste-Quadrate-Methode.

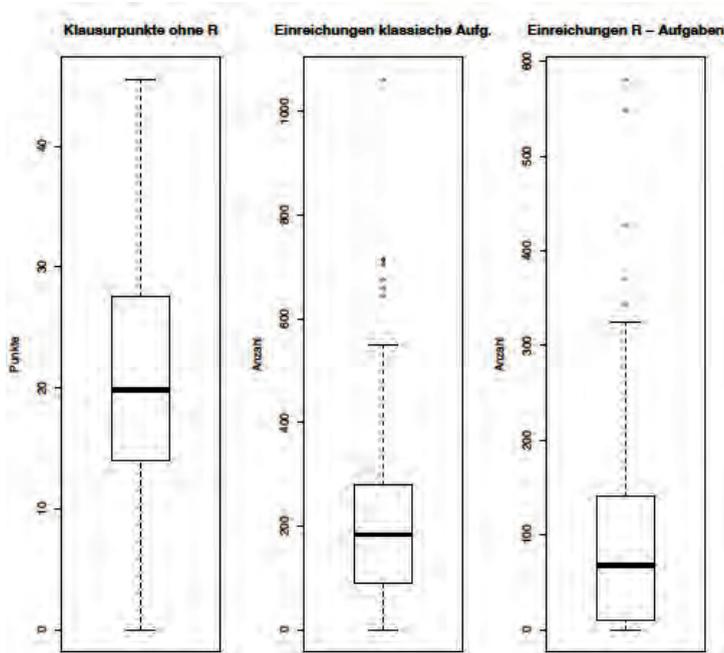


Abb. 5.1: Boxplots zur Veranschaulichung der Verteilung der für die Regression genutzten Variablen.

Abb. 5.2 zeigt ein Streudiagramm, welches die Klausurpunkte der Studierenden der Anzahl an Einreichungen in R-Übungsaufgaben gegenüberstellt. Eine beträchtliche Anzahl an Studierenden hat die R-Übungsaufgaben nicht oder nur sehr wenig bearbeitet. Diese schneiden zumeist auch in den klassischen Klausuraufgaben unterdurchschnittlich ab. Dafür geht eine gesteigerte Leistungsbereitschaft in R-Übungsaufgaben mit besseren Klausurergebnissen einher. Das Streudiagramm enthält zudem die Regressionsgerade der oben beschriebenen Regression. Dazu wurde in die Variable „Anzahl an Einreichungen in klassischen Aufgaben“ ihr Mittelwert eingesetzt.

Die geschätzten Regressionskoeffizienten sind Tab 5.2 zu entnehmen. Bemerkenswert ist der mehr als 3-mal größere Effekt der R-Übungsaufgaben im Vergleich mit den klassischen Übungsaufgaben auf die Klausuraufgaben. Möglicherweise stärken die Programmieraufgaben im Vorfeld das Verständnis der Materie deutlich stärker als die reproduktiven Rechenaufgaben. Für die Zukunft ist daher ein Ausbau des R-Bereiches

angedacht.

Wenn nun, siehe Tab. 5.3, die Klausurpunkte der R-Aufgabe (max. 8 Punkte) auf die gleichen Regressoren regressiert werden, wird auch hier ein positiver Effekt der R-Übungsaufgaben geschätzt. Der geschätzte Effekt der klassischen Übungsaufgaben ist hingegen nicht signifikant, da diese das Verständnis für Programmieraufgaben wohl nicht schärfen.

Koeffizient	Schätzung	Standardfehler	p-Wert
Konstante	15.53	0.79	~ 0
Anzahl R-Einreichungen	0.037	0.006	$8.52 \cdot 10^{-9}$
Anzahl Klass. Einreichungen	0.01	0.004	0.0175

Tab. 5.2: Geschätzte Koeffizienten, Standardfehler und p-Werte der Regression „Klausurpunkte ohne R“ auf „Anzahl Einreichungen in R-Aufgaben“ und „Anzahl Einreichungen in klassischen Aufgaben“.

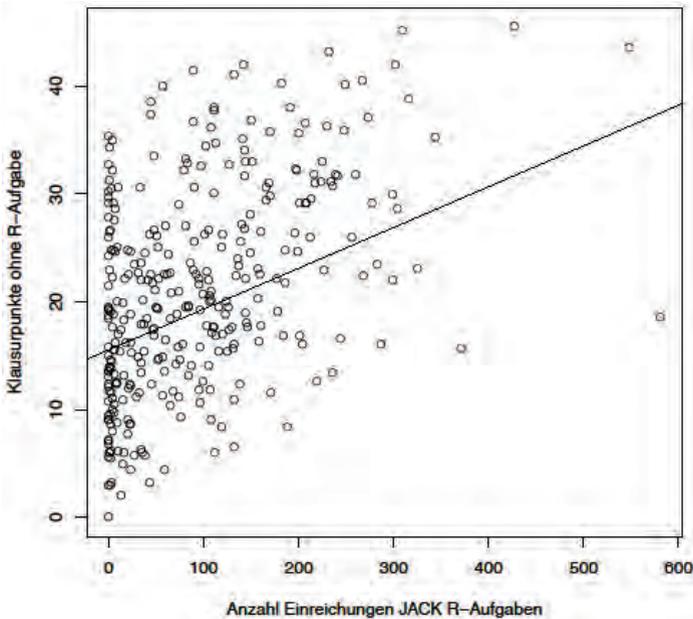


Abb. 5.2: Streudiagramm Klausurpunkte ohne R-Aufgabe gegen Einreichungen von R-Übungsaufgaben. Regressionsgerade der Regression Klausurpunkte ohne R gegen Einreichungen R-Aufgaben und Einreichungen klassische Aufgaben. In der Geraden wurde für die Kontrollvariable ihr Mittelwert eingesetzt.

Koeffizient	Schätzung	Standardfehler	p-Wert
Konstante	1.78	0.209	$5.08 \cdot 10^{-16}$
Anzahl R-Einreichungen	0.014	0.002	$1.35 \cdot 10^{-15}$
Anzahl Klass. Einreichungen	-0.001	0.001	0.346

Tab. 5.3: Geschätzte Koeffizienten, Standardfehler und p-Werte der Regression „Klausurpunkte der R-Aufgabe“ auf „Anzahl Einreichungen in R-Aufgaben“ und „Anzahl Einreichungen in klassischen Aufgaben“.

6 Fazit und Ausblick

Der eigens für die Statistikausbildung entwickelte JACK-R-Aufgabentyp wurde erstmals im WS 2016/17 genutzt, um den Studierenden der Vorlesung „Deskriptive Statistik“ die Anwendung der Statistiksoftware R zu vermitteln. Bei der Analyse der Nutzung dieses zusätzlichen Lehrangebots zeigte sich ein positiver Effekt von Programmier-Übungsaufgaben auf klassische Klausuraufgaben. Studierende, die das zusätzliche Lehrangebot nutzten, schnitten in der Klausur besser ab, als Studierende, die sich lediglich an den klassischen Aufgaben orientierten. Dies motiviert einen weiteren Ausbau dieses Konzepts auch in weiteren statistischen Veranstaltungen, um Studierende durchgehend besser zu fördern. Darüber hinaus soll die Förderung schwacher Studierender mit einem Ausbau des automatisierten individuellen Feedbacks verbessert werden. Zusätzlich arbeiten wir gerade an einer erweiterten Benutzeroberfläche in JACK, ähnlich dem Paket Rcmdr [FB17]. Damit sollen auch Studierende mit Nebenfach Statistik einen leichteren Zugang zu der Programmiersprache R finden.

Danksagung

Diese Arbeit wurde gefördert durch das Projekt „Professionalisierung für Vielfalt – Qualitätsoffensive Lehrerbildung“ des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01 JA 1610.

Literaturverzeichnis

- [Ba05] Basturk, R.: The Effectiveness of Computer-Assisted Instruction in Teaching Introductory Statistics. *Educational Society & Technology* 2/05, S. 170-178, 2005.
- [BB14] Buttner, E. H.; Black, A. N.: Assessment of the Effectiveness of an Online Learning System in Improving Student Test Performance. *Journal of Education for Business* 5/14, S. 248-256, 2014.

-
- [Ch07] Chance, B. et al.: The Role of Technology in Improving Learning of Statistics. *Technology Innovations in Statistics Education* 1/07, S. 1-26, 2007.
- [CM02] Chalmers, D.; McAusland, W. D. M.: Computer-assisted Assessment. Technical report, Glasgow Caledonian University, 2002.
- [Da17] DataCamp, datacamp.com, Stand: 21.03.2017.
- [FB17] Fox, J.; Bouchet-Valat, M.: Rcmdr: R Commander. R package version 2.3-2. 2017.
- [KJH16] Keuning, H.; Jeurig, J.; Heeren, B.: Towards a Systematic Review of Automated Feedback Generation for Programming Exercises. In: *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*, S. 41-46, 2016.
- [Ma17] Massing, T. et al.: E-Assessment Using Variable-Content Exercises in Mathematical Statistics. *Wiedereingereicht beim Journal of Statistics Education*, 2017.
- [Mc15] McNamara, A. A.: Bridging the Gap Between Tools for Learning and for Doing Statistics. Dissertation, 2015.
- [Me13] Means, B. et al.: The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. *Teachers College Record* 3/13, S. 1-47, 2013.
- [Mu15] Muenchen, R. A., The Popularity of Data Science Software, <http://r4stats.com/articles/popularity/>, Stand: 21.03.2017.
- [RC17] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, 2017.
- [Sc17] Schwinning, N. et al.: Towards digitalization of summative and formative assessment in academic teaching. *Proceedings of LaTiCE 2017*. Im Erscheinen, 2017.
- [So11] Sosa, G. W. et al.: Effectiveness of Computer-Assisted Instruction in Statistics: A Meta-Analysis. *Review of Educational Research* 1/11, S. 97-128, 2011.
- [St16] Striewe, M.: An architecture for modular grading and feedback generation for complex exercises. *Science of Computer Programming* 129/16, S. 35-47, 2016.
- [TL12] Tishkovskaya, S.; Lancaster, G. A.: Statistical education in the 21st Century: a review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education* 2/12, S. 195-272, 2012.
- [Un10] Universität Koblenz Landau, AG Ebert, <https://www.uni-koblenz-landau.de/de/koblenz/fb4/ist/rgebert/research/graph-technology/GReQL>, Stand: 17.03.2017.
- [Wi16] Wickham, H.: testthat: Unit Testing for R, <https://cran.r-project.org/web/packages/testthat/index.html>, Stand 20.03.2017.

Schulrechner wandern in die Cloud – Was bedeutet das für die unterschiedlichen Stakeholder?

Catrina Grella¹, Nils Karn², Jan Renz³, Christoph Meinel⁴

Abstract: Ausgehend von der momentanen Situation vieler Schulen in Deutschland zeigen wir vielfältige Chancen und Potenziale aber auch strukturelle, organisatorische sowie persönliche Bedenken hinsichtlich des Einsatzes von Cloud-Technologie und -Diensten in Schulen auf. Am Beispiel der derzeit unter Förderung durch das Bundesministerium für Bildung und Forschung (BMBF) entwickelten Schul-Cloud veranschaulichen wir, inwiefern digitale Angebote und ausgewählte Anwendungsszenarien zur Meisterung aktueller Herausforderungen an Schulen beitragen können – im Hinblick auf einen zeitgemäßen, fächerübergreifenden Unterricht, Kollaboration, adäquaten Vertretungsunterricht und Differenzierung.

Keywords: digitale Bildung, Schule, IT-Infrastruktur, Cloud, Bildungstechnologie.

1 Einleitung

Die digitale Transformation umfasst alle Lebensbereiche, auch den Bildungssektor: Die große Mehrheit der Lehrkräfte nutzen zur Vorbereitung ihres Unterrichts digitale Medien; 98 Prozent der Jugendlichen sind online [De14].

Nichtsdestotrotz stellt die informationstechnische Ausstattung viele Schulen vor große administrative und finanzielle Herausforderungen. Der Einsatz digitaler Medien in der Schule scheitert bislang an den unterschiedlichsten Faktoren: veraltete, nicht fachkundig administrierte Hard- und Software, unzureichende Internetanbindung, mangelhafte WLAN-Ausstattung, fehlende Nutzungs- sowie didaktische Konzepte, geringe Erfahrung der Lehrkräfte [He17] u. a. mit Bring your own device (BYOD), Sorge vor ungleichen Teilhabechancen aufgrund unterschiedlicher Technikausstattung, die Gefahr unerwünschten Verhaltens in sozialen Netzwerken oder Ablenkung bzw. Unterrichtsstörungen. Auch der „Bedarf, Bildungsinhalte und Kompetenzen zu vernetzen, spiegelt sich in den Lehrplänen der Länder, in der Ausbildung der Lehrkräfte und [...] der infrastrukturellen Ausstattung von Schulen in Deutschland oft noch nicht hinreichend wider“ [Bu16]. Im föderalistischen Bildungssystem, in dem Pluralismus ein Grundprinzip darstellt, werden zahlreiche Lehr- und Lerninhalte in „Silos“ verwahrt, deren Zugang teilweise von Zufälligkeiten abhängt. Im Rahmen des DigitalPakt

¹ Hasso-Plattner-Institut für Softwaresystemtechnik an der Universität Potsdam (HPI), Internet-Technologien und Systeme, Campus Griebnitzsee, 14482 Potsdam, catrina.grella@hpi.de

² HPI, Internet-Technologien und Systeme, Campus Griebnitzsee, 14482 Potsdam, nils.karn@hpi.de

³ HPI, Internet-Technologien und Systeme, Campus Griebnitzsee, 14482 Potsdam, jan.renz@hpi.de

⁴ HPI, Internet-Technologien und Systeme, Campus Griebnitzsee, 14482 Potsdam, christoph.meinel@hpi.de

zwischen Bund und Ländern sollen Schulen bundesweit mit der notwendigen Infrastruktur ausgestattet werden [Bu16], um zur Meisterung der digitalen Transformation sowie weiterer aktueller Herausforderungen, u. a. fächerverbindendem Unterricht, an Schulen beizutragen. Die Cloud für Schulen kann:

- eine ganzheitliche IT-Infrastruktur für eine unbegrenzte Anzahl an Schulen schaffen,
- flächendeckend die Nutzung neuester, professionell gewarteter IT-Systeme auch jenseits der Fächer Mathematik, Informatik, Naturwissenschaften und Technik (MINT) ermöglichen,
- ein breites Angebot an Inhalten schul- sowie bundeslandübergreifend leicht zugänglich machen,
- die Strukturierung sowie Koordination fächerverbindenden Unterrichts und die Kommunikation erleichtern sowie
- schulische und außerschulische Lernwelten besser miteinander vernetzen, um nachhaltige Lernprozesse zu erleichtern.

2 Bestehende Arbeiten & Potenzial der Schul-Cloud

Die Digitalisierung der schulischen Bildung steht im Spannungsfeld von Lernpraxis, Schulorganisation und Bildungspolitik. Während zahlreiche Lehrkräfte u. a. aufgrund empfundener Unsicherheiten noch vor dem Einsatz digitaler Medien zurückschrecken, kann technologiegestützter Unterricht vielen Schüler/innen zu Abwechslung sowie neuer Motivation verhelfen und auf die sogenannte „digitale Bürgerschaft“ [Of17] vorbereiten. Gute Beispiele u. a. aus Skandinavien zeigen, inwiefern Schule durch digitale Medien erfolgreich neu definiert werden kann, sofern Expertise, Mut und Durchhaltevermögen an den Tag gelegt werden. In Deutschland sind Ansätze digitaler Bildung indes oftmals noch lokal organisiert und Skaleneffekte werden kaum realisiert. Bestehende Angebote beschränken sich bislang insbesondere auf:

- Einzelfunktionalitäten wie den Austausch von Dateien,
- Einzelinitiativen engagierter Schulen, die für andere nicht nutzbar sind,
- komplex aufgebaute, unflexible und funktionsgetriebene Lernmanagementsysteme,
- proprietäre Lernplattformen, die an komplizierten Nutzungsszenarien ansetzen,
- Lösungen kommerzieller Anbieter, die aufgrund ihres restriktiven Lizenzmodells kaum skalierbar sind,
- Angebote internationaler Firmen, die nicht mit den in Deutschland geltenden Datenschutzrichtlinien vereinbar sind oder

- Ansätze, die auf einzelne Bundesländer beschränkt sind und kaum Synergiepotenziale nutzen.

Die Schul-Cloud soll die Aufwände für die Etablierung und den Betrieb von IT-Infrastrukturen an Schulen reduzieren, deren Leistungsfähigkeit erhöhen und möglichst viele Schüler/innen bundesweit von Bildungsinnovationen profitieren lassen. Gefördert vom Bundesministerium für Bildung und Forschung (BMBF) konzipiert das Hasso-Plattner-Institut (HPI) in Kooperation mit dem nationalen Excellence-Schulnetzwerk MINT-EC seit Herbst 2016 die Schul-Cloud. Durch sie können „bestehende Lösungen verschiedener Anbieter verknüpft und technisch angebunden werden. Lediglich Lösungen, die noch nicht existieren, werden neu entwickelt“ [Me17]. Im Sinne eines offenen, infrastrukturellen Elements kann die Schul-Cloud einer großen Gruppe von Nutzer/innen niedrigschwellig Zugang zu einem breiten Spektrum sowohl etablierter als auch alternativer digitaler Lernangebote ermöglichen – von Schulbuchverlagen über Startups, Open Educational Resources (OER) bis hin zu engagierten Einzelautor/innen. „Dabei sollen, soweit möglich, bestehende Standards verwendet werden“ [Me17], um eine leichte Anbindung unterschiedlicher Angebote in die Schul-Cloud zu ermöglichen. Je nach Fortschritt verwandter Initiativen zu zentralen Login-Diensten kann dieser Aspekt der Schul-Cloud durch solche Lösungen ergänzt werden.

„Die Schul-Cloud besteht aus mehreren Modulen, die als eigenständige und flexible Microservices implementiert“ [Me17] und über Schnittstellen angesprochen werden. Der modulare Aufbau erleichtert die Anbindung von Standardlösungen und Komponenten bestehender Systeme. „Die Schul-Cloud kann zum einen eigenständig genutzt werden, zum anderen können auch einzelne Komponenten herausgelöst verwendet werden. So ist es möglich, einzelne Dienste in bestehende Lernmanagement-Systeme einzubinden, um beispielsweise über die Schul-Cloud auffindbare Lernmaterialien auch in anderen Systemen, die an einigen Schulen bereits [...] eingeführt wurden, bereitzustellen“ [Me17].

Die Kernfunktionalitäten der Schul-Cloud umfassen zunächst:

- einen sicheren Cloud-Speicher zur Dateiablage (inkl. Virenscan und Vorschaubild),
- einen webbasierten Zugriff von unterschiedlichen Orten und Endgeräten aus,
- den Zugang zu einem großen Spektrum digitaler Lerninhalte unterschiedlicher Anbieter (Single Sign-on),
- die Bearbeitung und Abgabe von Hausaufgaben,
- den Kalender- und Benachrichtigungsdienst zum leichten Informationsaustausch und
- kollaboratives Arbeiten an gemeinsamen Aufgaben und Projekten.

3 Status Quo: Exzellente IT-Ausstattung an MINT-EC-Schulen?

Im November 2016 wurden die Leitungen der MINT-EC-Schulen anhand eines halbstandardisierten Fragebogens nach der aktuellen informationstechnischen Ausstattung ihrer Schule und ihren diesbezüglichen Anforderungen für die Zukunft befragt. Damit das konzipierte Angebot die Nutzung von Informationstechnologie in der Schule vereinfacht, muss es unmittelbar an die Bedarfe potenzieller Nutzer/innen anknüpfen und ihre Wünsche möglichst früh in die Softwareentwicklung einbeziehen – ganz im Sinn der Innovationsmethode Design Thinking [PMW09]. An der Erhebung beteiligten sich 67 MINT-EC-Schulen⁵. Für die frühe Pilotierung der Schul-Cloud ist MINT-EC aufgrund seiner bundesweiten Struktur sowie Technikaffinität und dem Exzellenzfokus, im Rahmen dessen das BMBF einzelne Schulen fördern darf, besonders geeignet. Langfristig ist die Schul-Cloud nicht notwendigerweise auf MINT-EC-Schulen begrenzt⁶. Die befragten Schulen stammen aus 14 Bundesländern sowie Österreich. Die Rücklaufquote lag bei 25 Prozent. Der Großteil der Teilnehmenden war zwischen 40 und 50 Jahren alt und männlich⁷.

Obwohl nahezu alle Befragten einen Schulserver haben (s. Abb. 1⁸) und es an zwei Drittel der Schulen (mindestens) einen Computerraum gibt, erscheinen diese eher als strukturelle Barriere als dass sie den Einsatz digitaler Medien erleichtern: Die bereitgestellte Hard- und Software veralten schnell und werden selten fachkundig administriert. Zudem stehen die separaten Computerräume häufig nur einzelnen Lehrkräften für bestimmte Fächer und begrenzte Zeiträume zur Verfügung. Diese können sich dadurch nicht längerfristig in diesem Raum einrichten und digitale Endgeräte spontan, bedarfsgerecht einsetzen. In etwa 60 Prozent der befragten Schulen gibt es (einzelne) Klassen, in denen alle Schüler/innen über einen Laptop oder ein Tablet verfügen und (teilweise) auch WLAN vorhanden ist. Eine Breitbandanbindung sowie leistungsfähiges, flächendeckendes WLAN in allen Klassenräumen ist für die flexible Nutzung digitaler Medien von höchster Priorität.

⁵ „exzellente MINT-Schulen, die bereits höchstes Niveau in der Qualität und Quantität des MINT-Unterrichts erreicht haben“ [Ve17]

⁶ Neben den MINT-EC-Schulen wurde auch die Perspektive von verwandten Projekten und Schulen, die weniger von Fördergeldern, -strukturen und Netzwerken profitieren, bereits in der Konzeptionsphase einbezogen, u. a. Grundschulen, integrierte Sekundarschulen, Mädchenschulen, freie Schulen, Berufsschulen und Volkshochschulen. Solchen interessierten Schulen kann vorbehaltlich der weiteren Förderung baldmöglichst auch ein Zugang zur Schul-Cloud eingerichtet werden.

⁷ „Frauen sind als Führungskräfte nach wie vor unterrepräsentiert – an Gymnasien stärker als an Grundschulen, an der Spitze der Schulleitung spürbarer als auf Stellvertreterposten“ [SB06].

⁸ Unter der Kategorie „Sonstiges“ sind die Nennungen von Beamern, interaktiven Tafeln, 3D-Druckern etc. subsummiert.

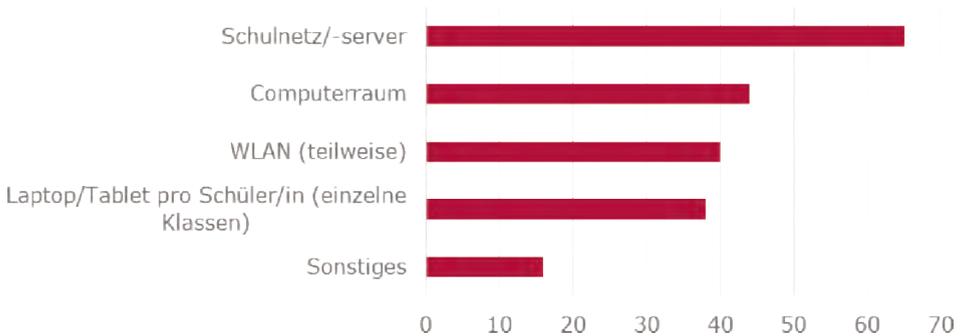


Abb. 1: IT-Ausstattung von MINT-EC-Schulen in Deutschland (N=67, Mehrfachantworten)

Die technische Administration übernehmen an mehr als 80 Prozent der befragten Schulen bisher engagierte Lehrkräfte (s. Abb. 2) – zusätzlich zu ihren Lehraufgaben. Weniger als 40 Prozent der befragten Schulen werden durch ein Dienstleistungsunternehmen unterstützt. An gut 20 Prozent der Schulen sind Mitarbeiter/innen aus der Verwaltung für die IT mitzuständig. An sieben befragten Schulen wird diese Aufgabe durch engagierte Schüler/innen übernommen. Die Zuständigkeiten von Schulträgern, kommunalen Medienzentren u. a. sind unter „Sonstiges“ zusammengefasst.



Abb. 2: Zuständigkeit für die technische Administration (N=67, Mehrfachantworten)

Über 60 Prozent der befragten Schulen administrieren ihre IT nach einem übergreifenden Konzept – zum Teil jedoch ohne zufriedenstellende Qualität. Fast die Hälfte der befragten Schulen geht mit auftretenden technischen Problemen situativ um.

Kritisch ist das nicht ausschließlich in Hinblick auf einen nutzerfreundlichen Einsatz digitaler Medien, für den eine professionelle Administration eine Grundvoraussetzung darstellt, sondern insbesondere auch in Bezug auf sicherheitsrelevante Aspekte.

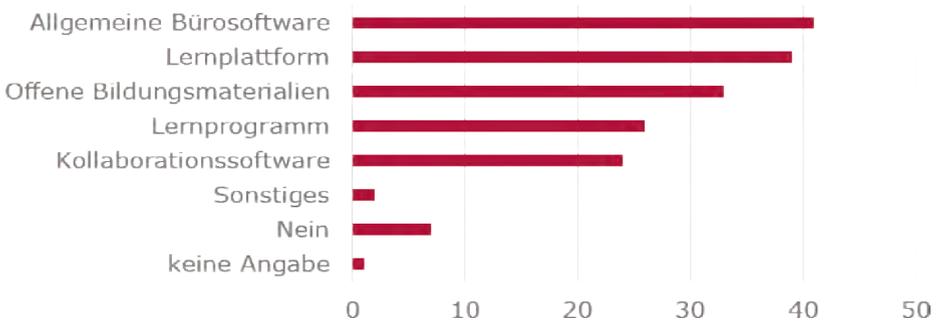


Abb. 3: Softwareeinsatz im Unterricht (N=67, Mehrfachantworten)

Analog zur heterogenen Ausstattung wird IT an den befragten Schulen derzeit sehr unterschiedlich eingesetzt. Obwohl MINT-EC-Schulen bereits bestimmte Grundvoraussetzungen im Bereich der Informationstechnik erfüllen [Ve17], setzen sieben der befragten Schulen bisher keinerlei Software im Unterricht ein (s. Abb. 3). An weiter vorangeschrittenen Schulen beschränkt sich der Einsatz digitaler Medien häufig auf allgemeine Bürosoftware (60 Prozent der befragten Schulen), u. a. Office-Programme und Lernplattformen (Moodle, Fronter, Lernsax uvm.). Gut die Hälfte der befragten Schulen verwenden offene Bildungsmaterialien wie digitale Aufgabenblätter und Präsentationen. Knapp 40 Prozent der Schulen nutzen bereits unterschiedlichste Lernprogramme (z. B. bettermarks, GeoGebra, schulbuchverlagsunterstützte Lernsoftware) sowie Vernetzungs-/ Kollaborationssoftware (u. a. lo-net2).

4 Was kann die Schul-Cloud leisten?

95 Prozent der befragten MINT-EC-Schulen wollen über die Schul-Cloud selbst erstellte sowie frei verfügbare Unterrichtsmaterialien für ihre Schüler/innen bereitstellen (s. Abb. 4). Der Anteil derer, die ihre eigens erstellten Unterrichtsmaterialien mit anderen Lehrkräfte teilen möchten, fällt mit 85 Prozent geringer aus. Für zwei Drittel der befragten Schulen ist es interessant, über die Schul-Cloud auf Unterrichtsmaterialien kommerzieller Anbieter zuzugreifen. Mehr als 90 Prozent möchten über die Cloud zudem Arbeitsergebnisse der Schüler/innen speichern. Gruppenarbeitsräume und Kommunikationstools erachten knapp drei Viertel der befragten Schulleitungen als wichtige Funktionalitäten. 80 Prozent der befragten Schulen scheint es sinnvoll, dass Lehrkräfte Funktionalitäten der Schul-Cloud (z. B. für Prüfungen) einschränken können. Weniger als die Hälfte der Befragten erkennen das Potenzial bei rechtlich einwandfreier Verwendung, über die Schul-Cloud Testtools und Learning Analytics zu nutzen.

Befragungsteilnehmer/innen im Alter von 50 Jahren oder älter scheinen in Hinblick auf die Nutzung von Testtools und Learning Analytics zurückhaltender als die jüngeren Befragten (50 Prozent der mindestens 50-Jährigen vs. zwei Drittel der Jüngeren). Als einen weiteren Bedarf meldeten einzelne Schulen u. a. ein sogenanntes Elternportal an.



Abb. 4: Benötigte Funktionalitäten der Schul-Cloud (N=67, Mehrfachantworten)

Das Spektrum benötigter fachspezifischer (Lern-) Software sowie Online-Angebote ist nahezu unbegrenzt. Eine detaillierte Wiedergabe der Erhebungsergebnisse würde den Rahmen dieses Papers sprengen. In Hinblick auf den Inhaltsbezug der erfassten Anwendungen und IT-Dienste ist jedoch festzuhalten, dass jenseits der MINT-Fächer insbesondere digitale Lernangebote in den Fremdsprachen, in Sport (z. B. zur Videoanalyse) sowie in Musik und Kunst an die Schul-Cloud angebunden werden sollen, um die Potenziale digitaler Medien fächerverbindend auszuschöpfen. Gemäß der Vielfalt und Schnellebigkeit digitaler Lernsoftware kann eine Art dynamischer „Lern-Store“⁹ eine niedrigschwellige, flexible Nutzung ermöglichen.

5 Anwendungsszenarien der Schul-Cloud

Unterschiedlichste Nutzungsmodelle der Schul-Cloud sind denkbar. Mehr als 90 Prozent der Befragten sehen Mehrwerte insbesondere in Bezug auf die in Lernprozessen elementare Wiederholung und Auffrischung von Inhalten sowie Differenzierung [PL12], d. h. der individuellen Förderung durch spezifische Lernangebote (z. B. bezüglich Lerntempo und -typ). Trotz des zunehmenden Anteils an Ganztagschulen erachten fast 80 Prozent der befragten Schulleitungen die Cloud als hilfreich für Hausaufgaben. Die große Relevanz der Vernetzung unterschiedlicher Lernorte (sowohl in Hinblick auf die Verbindung von Schule und dem Nachmittagsmarkt sowie formalem, non-formalem und

⁹ analog zum App-Store von Apple oder Play-Store von Google

informellem Lernen) spiegelt sich im Bedarf an Blended Learning-Szenarien wieder, d. h. der Kombination aus Präsenzunterricht und digital-unterstütztem Lernen, z. B. die strukturierte Vor- oder Nachbereitung von Unterricht durch webbasierte Angebote.

Ferner sehen die befragten Schulen große Chancen über die Cloud projektorientiertes [Sc15a]¹⁰ und wissenschaftspropädeutisches¹¹ [Bo12] sowie eigenverantwortliches Arbeiten¹² und forschendes Lernen¹³ zu fördern. Komplexe Themen, die an den Erfahrungsbereich der Schüler/innen anknüpfen, werden über die Schul-Cloud anhand von innovativen Lerninhalten fächerverbindend nutzbar. Auch weitere bildungsnahe Akteure sowie Projekte können über die Cloud interessanten Content und didaktische Materialien teilen. Beispielhaft sei auf BeeBIT hingewiesen, „ein in Deutschland ansässiger Verein, der sich auf Bienenforschung spezialisiert hat, indem er ein Netzwerk aus mit Sensoren erweiterten Bienenstöcken über ganz Europa verteilt, aufbaut“ [Be16]. Die aus mehreren elektronischen Bienenstöcken generierten und zahlreichen Schüler/innen bereitgestellten Echtzeitdaten ermöglichen bundeslandübergreifende Forschungsprojekte an der Schnittstelle zwischen Biologie, Geographie und Mathematik. Durch ihren Anwendungs- sowie haptischen Bezug unterstützen solche Angebote im Sinne der Reformpädagogik¹⁴ beim Aufbau digitaler Kompetenzen.

Der Austausch digitaler Materialien im Kollegium und innerhalb der Schülerschaft ist insbesondere auch bei (krankheitsbedingter) Abwesenheit hilfreich: z. B. können sogenannte Vertretungspakete von zu Hause aus geschnürt sowie zur Anwendung durch eine Vertretungslehrkraft freigegeben werden. Auch Schüler/innen, die einmal nicht zur Schule kommen können, haben online in Echtzeit Zugriff auf Unterrichtsinhalte und Hausaufgaben. Ausfälle sowie Nachholbedarfe können dadurch minimiert und die Rückkehr in den laufenden Schulbetrieb erleichtert werden.

Die Schul-Cloud kann ferner die Portfolio-Arbeit¹⁵, Dokumentation individueller Lernleistungen und Entwicklung von digitalen bzw. Computer und

¹⁰ Das breite Spektrum digital verfügbarer Informationen kann projektorientiertes Lernen in der Schule maßgeblich erleichtern. „Projektorientiertes Lernen wird besonders aus konstruktivistischer Perspektive befürwortet, da davon ausgegangen wird, dass es den Erwerb bedeutungsvollen Wissens, das sich die Lernenden in selbstständiger Weise aneignen, fördern sollte“ [Sc15a].

¹¹ Das Kennenlernen und selbstständige, problemlösende Anwenden wissenschaftlicher Methoden und gruppendynamischer Arbeitsprozesse, insbesondere auch anhand digitaler Medien (z. B. allein schon die reflektierte Internetrecherche), stellen wichtige Voraussetzungen für das Leben in der Wissensgesellschaft dar [Bo12].

¹² Digitale „neue Medien erscheinen vielen Bildungsverantwortlichen, Forschern und Praktikern als ideale Instrumente für eigenverantwortliches und selbstgesteuertes Lernen [...]; hohe Selbststeuerung bedeutet jedoch nicht immer auch optimalen Lernerfolg“ [FMT09].

¹³ „Digitale Medien können beispielsweise die soziale Situiertheit in die wissenschaftliche Community über Wissenschaftlerblogs oder die Teilnahme an Tagungen im ‚second life‘ abbilden“ [DRH11].

¹⁴ für weiterführende Informationen siehe auch [Sc15b]

¹⁵ „Portfolios können auf unterschiedliche Weise technisch realisiert werden: Die Formulare werden *handschriftlich* oder *elektronisch* ausgefüllt, ‚Portfolio-Strukturen‘ können auf dem Computer zur Verfügung gestellt oder dort von den Besitzern selbst eingerichtet werden. Sie lassen sich auf einem persönlichen, externen Speichermedium (z. B. USB-Stick) ablegen, als eigenen Ordner in einer schulinternen, netzgestützten Arbeitsplattform oder im Internet (e-Portfolios) [Ni10].

informationsbezogenen Kompetenzen [Bo13] unterstützen und Schüler/innen die Lerninhalte und Werkzeuge bieten, die ihren jeweiligen Bedürfnissen entsprechen. Über die erläuterten Potenziale im pädagogischen Bereich hinausgehend kann die Schul-Cloud, z. B. anhand von Umfragen und Tools der formativen Evaluation aus Sicht der befragten Schulleitungen auch diverse organisatorische Belange im schulischen Umfeld erleichtern.

6 Einführung der Schul-Cloud in 27 Pilotschulen

Auch wenn die Aufgeschlossenheit der teilnehmenden Pilotschulen in Hinblick auf mögliche Veränderungen von Lehren, Lernen, Schule und der Rolle von Lehrkräften durch den Einsatz digitaler Medien vorausgesetzt wird, ist die Heterogenität zwischen und innerhalb der Schulen bei ihrer schrittweisen Annäherung an die Schul-Cloud zu berücksichtigen. Eine von digitalen Technologien mitgeprägte Lernumgebung, ist in einem komplexen Gefüge aus Technologie, Pädagogik und Inhalten verortet. In Anlehnung an das Konzept „Technological Pedagogical Content Knowledge“ (TPACK) sind für die effektive Nutzung digitaler Medien in der Lehre spezifische Kenntnisse der Lehrkräfte hilfreich [MK06]:

- Technologiewissen,
- Inhaltswissen,
- pädagogische Kenntnisse,
- pädagogisch-inhaltliche Kenntnisse (u. a. Fachdidaktik),
- technologiespezifisches Inhaltswissen (z. B. technologische Möglichkeiten zur Vermittlung von Inhalten),
- technologisch-pädagogisches Wissen (Medienpädagogik) und
- technologisch-pädagogisches Inhaltswissen (Zusammenspiel von Technologie, Pädagogik und Inhalten im Lehr-Lern-Geschehen).

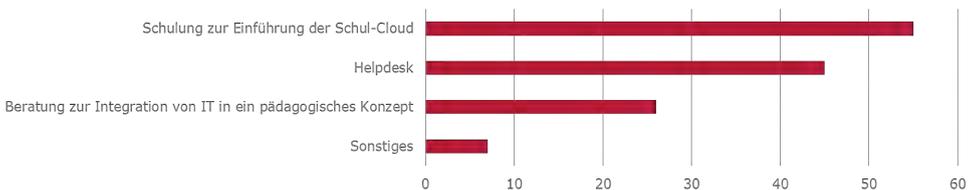


Abb. 5: Unterstützungsbedarfe seitens der befragten Schulen (N=67, Mehrfachantworten)

Die Digitalisierung der Bildung kann nur zum Erfolg führen, wenn neben der technischen Infrastruktur digitale Inhalte, didaktische Modelle, die Lehrkräfteweiterbildung und Finanzierung berücksichtigt werden [Bi16]. Mehr als 80

Prozent der befragten Schulen haben Interesse an Schulungen sowie Tutorials zur Schul-Cloud (s. Abb. 5). Zwei Drittel halten die Einrichtung eines sogenannten Helpdesks für unabdingbar. Unterstützung bei der Integration von IT in ein pädagogisches Gesamtkonzept sehen gut ein Drittel der befragten Schulen. Als sehr hilfreich erachten sie einen engen Austausch u. a. zu Erfahrungen und guten Praxisbeispielen im Rahmen eines (regionalen) Netzwerks sowie Informationen zu technischen Mindestanforderungen (Breitband, WLAN). In drei Arbeitsgruppen zu den Themen „Schulleitung und -organisation“, „Nutzung der Schul-Cloud im Unterricht“ sowie „Einsatz von IT in der Schule“ arbeiten die teilnehmenden Schulen während der gesamten Pilotphase an ausgewählten Aspekten, offenen Fragen und Konzepten der praktischen Anwendung der Schul-Cloud. Die erarbeiteten Ergebnisse werden in Form einer Handreichung für weitere interessierte Schulen aufbereitet, während die Mitglieder der Arbeitsgruppen die Rolle als Multiplikator/innen übernehmen – ähnlich wie eine „Referenzschule für Medienbildung“ [Ru12].

Dem Modell „SAMR – Substitution, Augmentation, Modification, Redefinition“ [Pe03] zufolge ist es erfolversprechend, Lehrkräfte schrittweise an die Nutzung digitaler Lehrmittel über die Schul-Cloud heranzuführen. Zunächst können auf die bereits eingesetzten Lehrmaterialien abgestimmte digitale Lernangebote in der gleichen Funktion eingesetzt werden wie bisherige Inhalte. Die Schul-Cloud wird dadurch niedrigschwellig in die gewohnte Unterrichtsvorbereitung eingebunden. Im zweiten Schritt werden erweiterte, z. B. interaktive, Funktionalitäten genutzt, die ausschließlich analoge Materialien nicht oder nur sehr umständlich ermöglichen. Im dritten Schritt können Lehrmaterialien anhand von digitalen Medien so umgestaltet werden, dass ihre Mehrwerte unmittelbar erlebt werden¹⁶. Auf der höchsten Stufe sind neuartige Lehrmaterialien zu verorten, z. B. Visualisierungen, die ohne digitale Medien kaum einsetzbar sind.

7 Fazit & Ausblick

Über die Schul-Cloud werden erstmalig Erfahrungen mit einer speziell für den schulischen Bereich entwickelten Cloud-Lösung in einem bundesweiten Kontext gesammelt. Hierbei werden nicht nur Potenziale von Cloud-Strukturen und -Diensten im schulischen Bereich gehoben. Vielmehr können sich über die Schul-Cloud komfortable Möglichkeiten für die schulinterne sowie -externe Vernetzung und bundeslandübergreifende Synergien etablieren. Vielfältige Lerngelegenheiten entstehen für den frühzeitigen Erwerb und Ausbau digitaler Kompetenzen (explizit und implizit). Schulbuchverlage und alternative Über die Schul-Cloud testen Content-Anbieter innovative Lehr- und Lernangebote auf ihre Anwendbarkeit und machen sie im schulischen Kontext bekannt. Attraktive Lizenzmodelle, innovative Verteilungsszenarien sowie -kanäle für digitale Angebote können sich auf diesem Weg auch in Fächern

¹⁶ ähnlich wie in Hinblick auf das sogenannte „WeQ“ beim kollaborativen Arbeiten [DM15]

entwickeln, in denen sich die Inhalte nur langsam verändern.

Zukünftig werden im Forschungsprojekt „Schul-Cloud“ u. a. auch die folgenden Fragen adressiert: Wie kann die aus technischer Sicht leicht zu ermöglichende Skalierung in die Fläche getragen werden? Welche infrastrukturellen Rahmenbedingungen können weiteren Schulen an die Hand gegeben werden, sodass die Schul-Cloud von möglichst vielen Schüler/innen genutzt werden kann? Wie wirkt sich die Vernetzung unterschiedlicher Lernorte durch digitale Medien auf das Lehr- und Lerngeschehen aus?

Literaturverzeichnis

- [Be16] BeeBit e.V.: BeeBIT. Würzburg, 2016. <https://beebit.de/de>, 24.3.2017.
- [Bi16] Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V.: Digitale Bildung – Handlungsempfehlungen für den Bildungsstandort Deutschland. Positionspapier. Berlin, 2016.
- [Bo12] Boggasch, Mirjam: Wissenschaftspropädeutik in der Schule – Musik und Literatur als wissenschaftspropädeutisches Seminar in der gymnasialen Oberstufe. Dissertation zur Erlangung des Doktorgrades der Fakultät „Musik“ der Universität der Künste, Mannheim/Neckarau.
- [Bo13] Bos, Wilfried et al.: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern in der 8. Jahrgangsstufe im internationalen Vergleich. Waxmann, Münster, New York, 2014.
- [Bu16] Bundesministerium für Bildung und Forschung: Bundesoffensive für die digitale Wissensgesellschaft. Strategie des Bundesministeriums für Bildung und Forschung, Berlin, 2016.
- [De14] Deutsches Institut für Vertrauen und Sicherheit im Internet (DIVSI): DIVSI U25-Studie. Kinder, Jugendliche und junge Erwachsene in der digitalen Welt. Hamburg, 2014.
- [DM15] Dräger, Jörg; Müller-Eiselt, Ralph: Die digitale Bildungsrevolution. Der radikale Wandel des Lernens und wie wir ihn gestalten können. Deutsche Verlags-Anstalt, München, 2015.
- [DRH11] Dürnberger, Hannah; Reim, Bettina; Hofhues, Sandra: Forschendes Lernen: konzeptuelle Grundlagen und Potenziale digitaler Medien. In: Köhler, Thomas; Neumann, Jörg (Hg.): Wissensgemeinschaften. Digitale Medien – Öffnung und Offenheit in Forschung und Lehre. Münster; New York; München; Berlin: Waxmann 2011: 209-219. – (Medien in der Wissenschaft; 60).
- [FMT09] Fischer, Frank; Mandl, Heinz; Todorova, Albena: Lehren und Lernen mit neuen Medien. In: Tippelt, Rudolf; Schmidt, Bernhard (Hg.): Handbuch Bildungsforschung, 2., überarbeitete und erweiterte Auflage, VS Verlag für Sozialwissenschaften | GWV Fachverlage GmbH, Wiesbaden, 2009.
- [He17] Heise online: Lehrerverband beklagt mangelhafte digitale Ausstattung an Schulen. 2017, <https://www.heise.de/newsticker/meldung/Lehrerverband-beklagt-mangelhafte->

- digitale-Ausstattung-an-Schulen-3663645.html, 24.3.2017.
- [Me17] Meinel, Christoph; Renz, Jan; Grella, Catrina; Karn, Nils; Hagedorn, Christiane: Die Cloud für Schulen in Deutschland: Konzept und Pilotierung der Schul-Cloud. Technische Berichte Nr. 116, 2017.
- [MK06] Mishra, Punya; Koehler, Matthew J.: Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. Teachers College Record Volume 108/2006, 1017-1054, 2006.
- [Ni10] Niedersächsisches Landesamt für Lehrerbildung und Schulentwicklung (NiLS) (Hg.): Leitfaden Portfolio: Medienkompetenz, 2. erweiterte Auflage: 500 – Hildesheim, 2010.
- [Of17] Office of Educational Technology: Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update, Version 2.0. 2017, <https://tech.ed.gov/files/2017/01/NETP17.pdf>, 24.3.2017.
- [Pe03] Puentedura, Ruben R.: A Matrix Model for Designing and Assessing Network-Enhanced Courses. 2003, 216.92.11.199/resources/matrixmodel/puentedura_model.pdf, 21.3.2017.
- [PL12] Paradies, Liane; Linser, Hans J.: Differenzieren im Unterricht, 6. Auflage, Cornelsen Verlag-Scriptor GmbH & Co. KG, Berlin, 2012.
- [PMW09] Plattner, Hasso; Meinel, Christoph; Weinberg, Ulrich: Design Thinking – Innovation lernen – Ideenwelten öffnen. mi-Wirtschaftsbuch, München, 2009.
- [Ru12] Ruppert, André: Referenzschule für Medienbildung. München, 2012, https://www.mebis.bayern.de/wp-content/uploads/sites/2/2015/05/Leitfaden_Referenzschulen_2012.pdf, 24.3.2017.
- [SB06] Scheuring, Andrea; Burkhardt, Anke: Schullaufbahn und Geschlecht. Beschäftigungssituation und Karriereverlauf an allgemeinbildenden Schulen in Deutschland aus gleichstellungspolitischer Sicht, HoF-Arbeitsberichte 4, 2006.
- [Sc15a] Schaumburg, Heike im Auftrag der Bertelsmann Stiftung: Chancen und Risiken digitaler Medien in der Schule. Medienpädagogische und -didaktische Perspektiven, Gütersloh, 2015.
- [Sc15b] Schubert, Maike; Grams, Stefanie; Röschelsen, Lars; Hubert, Olaf: BYOD und selbstgesteuertes Lernen. Das Projekt *MediaMatters!* An der Freiherr-vom-Stein-Schule Neumünster. In: Computer + Unterricht 25 (2015), Heft 99: 31-32.
- [Ve17] Verein mathematisch-naturwissenschaftlicher Excellence-Center an Schulen e.V. MINT-EC – Das nationale Excellence-Schulnetzwerk, <https://www.mint-ec.de/>, 09.03.2017.

A Crowdsourcing-based Learning Approach to activate Active Learning

Agnes Koschmider¹, Mario Schaarschmidt²

Abstract: Usually students consume learning material and write an exam at the end of the lecture. Such a process follows a summative learning pattern, which can be considered a standard approach at universities. Studies in educational theory indicate, however, that active involvement – instead of passive consumption – should be fostered in learning since active learning proved to be superior to passive learning. To benefit from active learning arrangements, we implemented an active involvement of students into the exam preparation for an introduction to Information Systems course at the University of Cologne. Students were asked to design exercises and provide solutions to selected topics. Subsequently, they received feedback to their submissions, which supports the self-assessment on the subject. An empirical evaluation shows general agreement for such active involvement of students and also indicates that students participating in the task creation are more likely to pass an exam than students denying the participation. This paper presents our crowdsourcing-based learning approach and discusses challenges for its implementation.

Keywords: active learning, learning map, crowdsourcing, information systems education

1 Introduction

Paper-written exams are still the first choice at universities. Due to a reduced effort for the marking of exams, electronically-based examinations (e-exams) are becoming increasingly common [Wi16]. Although a significantly reduced time for marking of e-exams favors them against paper-written exams, still, e-exams have to cope with the challenge that they do not foster individual capabilities of learners. Looking at the process of preparation of e-exams from a role perspective (student, teacher), a variety of activities are undertaken by the lecturer him/herself, who thus takes an active role. The students remain passive, which means that only their passive learning is addressed. According to the Center of Research and Learning (2016) active learning “*is a process whereby students engage in activities, such as reading, writing, discussion, or problem solving that promote analysis, synthesis, and evaluation of class content*”. Studies comparing active against passive learning show that active learning outperforms passive learning in various dimensions such as learning success [We12].

A novel approach of student involvement has been tried at the University of Cologne as a specificity of active learning. Students received the opportunity of an exam bonus for the lecture “Foundations of Information Systems” in the summer term 2016. Their task

¹ Institut AIFB, Karlsruher Institut für Technologie, 76128 Karlsruhe, agnes.koschmider@kit.edu

² ZIHET, Universität Koblenz-Landau, mario.schaarschmidt@uni-koblenz.de

was to design exercises *and* solutions to selected topics and to assign complexity scores to each exercise. One lecturer or teaching assistant was responsible for the marking of a single exercise and accompanying solution proposal and for communicating feedback to students. Peers could also review the submissions of others. About 100 students participated and submitted five exercises and solutions to topics such as Event-Driven Process Chain (EPC), Python, SQL, UML, ERM and HTML. An empirical evaluation that we conducted at the end of the term supports our crowdsourcing-based learning approach and shows that students that participated were more likely to pass an exam than students denying participation. Particularly, the following reasons justify this result from the empirical study:

- The continuous design of exercises and solutions in the semester is a representative of formative learning, which fosters active learning since peers build a map of the learning material already when designing exercises and solutions.
- Lectures received feedback concerning lacks of understandability that pop-upped within the evaluation of exercises and solutions. After the marking of all tasks it became evident that particular topics were not well understood. Lectures can use this feedback in order to recap topics and to improve the teaching [LN05]. Thus, misunderstanding in the learning material is clarified along the course.

Beside the positive effect of passing an exam, the approach of exercise and solution design by students can be used as a foundation to implement individual and individualized e-exams. The large set of exercises forms a repository and is suitable to provide e-exams at individual time and considering individual preferences of students with respect to visualization of exam questions. Moreover, the evaluation shed light on various factors for future intention to participate again in such a bonus program, thus suggesting aspects on how to successfully design such a crowdsourcing-based approach.

This paper of piloting a crowdsourcing-based learning approach is organized as follows. The next section relates our approach to existing approaches in the areas of crowdsourcing and active learning. Section 3 compares the as-is and to-be process of exam preparation and thus demonstrates the benefits of a crowdsourcing-based learning approach. Section 4 presents the results of an empirical evaluation study in the summer term 2016, and Section 5 discusses our lessons learnt. The paper concludes in Section 6 with a discussion on future directions to fully implement such an approach.

2 Related Work

Crowdsourcing refers to spreading tasks to a mostly unknown workforce of professionals or everyday people [Ho06]. The concept has been widely adopted such as for idea generation [Le09]. In line with these applications, crowdsourcing-based learning is related to distributing learning tasks and a suitable approach to support e-assessment and peer evaluation, which are two well established learning instruments. E-assessment

refers to formative, self-assessing of online exams where the lecturers provide the (exam) question pool [De09]. Open question pools of other universities might be integrated in order to broaden the view of the learning material [LN11]. However, the active learning effects are not fully exploited in e-assessment. Within a crowdsourcing-based learning approach, the learners design exercises and solutions and also revise the quality of questions of other peers. Consequently, with such a learning approach their active learning is addressed.

Mutual feedback by students is referred by peer evaluations, which can be considered as a specialization of e-assessment. Students mutually evaluate the achievements of their peers. Numerous empirical studies showed positive effects for this learning instrument [Bo01, LL15]. This finding also resonates with [Sc15], who recommends peer reviewing for seminar papers. A crowdsourcing-based approach goes beyond peer assessment. Learners are requested to design exam exercises and thus to build a map of the learning material [No10]. Their task is to assess the quality of their peers. Moreover, they also have to indicate and evaluate complexity scores of exercises and they must familiarize themselves with representation of exam tasks, which might be different to the own preferences (e.g., visual vs. text). Approaches in favor of using crowdsourcing in the learning process can be found in the literature and they support our approach. For instance, [We12] recommends crowdsourcing in combination with personalized online education in order to reach full potentials of online education. [An11] points to positive effects of crowdsourcing for judging of answers by peers. [CM12] found out that crowdsourcing can offer additional richness for accreditation and assessment.

To sum up, crowdsourcing in learning is in its infancy particularly when it comes to practical use in learning. Initial empirical studies, however, indicate positive effects compared to conventional learning. Thus, these studies justify our research of a crowdsourcing-based learning approach. The initial idea of our crowdsourcing-based learning approach thus without any presentation of evaluation, discussions and lessons learnt has been published within a short paper [KB16].

3 As-is and to-be process of exam preparation

To elucidate the benefits of the crowdsourcing-based learning approach we now compare it with the common as-is process of exam preparation. The as-is process starts with the preparation of exam questions by the lecturer based on a selected set of topics. Subsequently the questions are composed to an exam where quality assurance (i.e., acknowledging the correctness of the way exam questions are proposed along with ensuring suitable solutions) is subject to lecturers. At a predefined time, the exams are written either on paper or at a computer-based system and the exam is supervised by the lecturer or research assistants. Next, the exams are marked either by a system (in case of a multiple-choice-based e-exam) or by hand (in case of paper-written exams). Finally, the results are published.

Against this “conventional” exam process, we suggest the following process (*to-be*), which is inspired by crowdsourcing [Ho06]. Instead of writing exams that are predefined by lecturers, students provide exercises and solutions for exams. They indicate levels of difficulty and scores for the questions. The evaluation of exercises and solutions is mutually conducted by the students. The students also mutually benchmark, validate the scores of the exercises and solutions and improve exercises based on a star rating, which is common to them (e.g., classification for hotels, restaurants or services in the internet). The benchmark can be done using a system being equipped with corresponding features. The final quality of the exam questions and solutions, however, is still subject to the lecturer. He/she decides if exercises are still not mature enough or not appropriate as exam questions and drops them out of the system. When an exam should be conducted, then the system composes an exam out of the students’ exercises. Comparing the as-is against the to-be process it becomes evident that learners are more involved in the exam process in the crowdsourcing-based learning approach. The to-be process has been partially implemented within the lecture “Foundations of Information Systems” in the summer term 2016. A bonus of at least 15 points (17% of the total exam score) was granted to students if they submitted exercises and solutions to 5 out of 6 topics such as EPC, Python, SQL, UML, ERM and HTML. Around 120 to 150 students attended the lecture on a regularly basis. For each task (i.e., each round of submitting questions and solution), students received a document summarizing the task, which also highlighted associated learning targets. In particular, for EPCs, for example, we stated that “*students should already have understood the syntax and semantics of the Event-Driven Process Chain and have to apply the modeling notation correctly.*” This learning target excluded exercises towards the syntax of EPC such as “Right or wrong: an EPC consists of rectangles”. After this initial clarification, example exercises, solutions, and possible difficulty scores were communicated to the students as well as a tool recommendation. The students had one week of time for submission. One lecturer or teaching assistant was in charge of marking a topic in order to detect cheating and duplicates in the question pool. Examples of very good and comparatively bad question-solution pairs were discussed in the next session. Individual feedback could be provided upon request. In addition, topics were recapped where a lack of understandability was identified.

To validate our crowdsourcing-based approach demonstrated on the process of exam preparation the next section presents results of an empirical study.

4 Evaluation

The evaluation of the crowdsourcing-based learning approach is split into two separate parts. Part 1 involves the analysis of students’ success in both the bonus program as well as the final exam as such. Part 2 consists of a post-exam survey among students to identify drivers and barriers of participation and future intention to participate in such a program.

Concerning part 1, we analyzed exam and bonus program results. For reasons of anonymity, we only compared exam and bonus program results, but did not merge them with the survey in part 2 or any other demographic (e.g., age, semester). Officially, 227 students were registered to the course, of which approximately 120-150 participated on a regularly basis (based on simple counting by the lecturer). Students could choose among two exam dates, one two weeks after the last session and the other at the beginning of the next semester. Our analysis is limited to students who participated in the first exam. Ninety students took part in this exam. Of these students, 83 participated in the bonus program, meaning that they submitted at least one out of five exercises. The students could receive 15 bonus points in total. On a descriptive basis, the following aspects are notable (without providing theoretical explanations):

- Ten students did not pass the exam, of which seven did not participate in the bonus program.
- Students who scored 1.0 or 1.3 in the exam (the two best possible grades in the German system) received 13.28 bonus points (SD = .3) on average.
- The average grade of those students who had received 14 or 15 bonus points, got a median grade of 1.7 (Note: Possible grades are 1.0, 1.3, 1.7, 2.0, 2.3, 2.7, ... 5.0, i.e., not passed)
- Students who would not have passed the exam without applying bonus points got 10.4 bonus points on average.

Part 2 of our evaluation is based on a survey among course participants. First, shortly after the course has ended, we discussed the perceptions concerning the bonus concept as such with a group of selected students as well as course tutors. Their perceptions helped to identify questions of concern and helped to abandon questions that were not perceived as relevant by the target group. We also discussed the entire bonus program-approach as well as the evaluation with experts in didactics and learning.

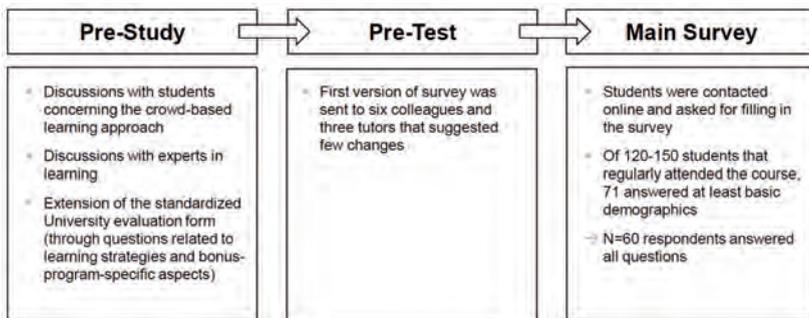


Fig. 1: Process of survey-based evaluation

Based on these insights, we developed a first version of a questionnaire, which consisted of university-wide standardized questions pertaining to the course and the lecturer, enriched by questions related to the bonus program and exam. Finally, students were contacted online through the course management system and were asked to complete the

questionnaire. This invitation was sent out three weeks after the first exam was held. We sent out one reminder one week later. Of the 227 students registered in the course system, 84 started to answer questions, but only 60 completed the entire survey. The entire process of Part 2 of our evaluation is depicted in Figure 3.

Among the 60 students who completely filled in the questionnaire, 34 (56.7%) were women, which reflects the course structure quite well (Note: 95% of the students were enrolled in business administration). Respondents' average age was 22 years and the mean semester they studied in was 4.23. We also asked about their average grades received prior to taking the course. Respondents indicated to have an average grade of 1.67 (in a German system were grades start with 1 [best] until 5 [not passed]). We also assessed the effort a student invested in a) preparing the exam and b) preparing exercises for the bonus program. Respondents indicated to have invested a total of about 60 hours for exam preparation, and about 3.5 hours per exercise they submitted. Table 1 provides an overview on descriptive aspects of the sample.

	#	%	Mean	SD
Gender				
Male	26	43.3%		
Female	34	56.7%		
Age			22.18	2.30
Semester			4.23	1.94
Average grade (self-rated)			1.67	.94
Effort for exam preparation (in hours; total)*			60.68	40.96
Effort for bonus program participation (in hours, per exercise)			3.52	2.62

*Only $N=40$ students who attended the first exam answered this question.

Tab. 1: Descriptive evaluation results

Another reliable indicator of this crowdsourcing-based learning approach's success (apart from the ratio of failed exams) is the students' future intention to participate in the program. The survey therefore also contained an item that asked whether or not students would participate again in such a bonus program. Of the 60 respondents, 51 agreed to be willing to participate again, while only nine indicated to not be willing to participate again. While the overall number of 85% agreement on future participation seems promising, we were also interested in what drives or impedes future intention. Based on discussions mentioned above, we decided to focus on a distinction between bonus program specific aspects of future intention and aspects that were related to students' learning strategies and achievement goal orientations. For goal orientation, we relied on the conceptualization by Ames and Archer [AA88], who distinguish mastery goal orientation from performance goal orientation. Students that favor a mastery goal performance are driven by the willingness to develop new skills and abilities. They value the process of learning itself regardless of the final outcome. With a performance goal

orientation, in contrast, students value normative high outcomes such as achieving success with little effort or outperforming others. [AA88] differentiate both forms of achievement goals in terms of eight dimensions. We adopt three of these eight dimensions based on the applicability to our research context: aspects of success, reasons for satisfaction, and focus of attention. For each facet, we adapted one item for both achievement goal orientations (see Table 2). In addition, we integrated two items on deep learning strategies, also inspired by Ames and Archer [AA88].

Question	Factor 1	Factor 2	Factor 3
1 I rate success in my studies as the reception of good grades. [PER1]	.829		
2 I am satisfied with my studies when I am better than others. [PER2]	.864		
4 I rate success in my studies as my personal improvements. [MAS1]		.526	
5 I am satisfied with my studies when I have worked hard [MAS2]		.778	
6 I focus my attention primarily to my own learning progress. [MAS3]		.656	
7 During the course “_____”, I tried to pursue a deep learning strategy [LS1]			.852
8 I set my own goals for the course “_____”.			.796

*Note: Item 3 was not included in this analysis, because it revealed a low factor loading of below .5. Values below .4 are not displayed

Tab. 2: Factor analysis on goal orientation and learnings strategy

To analyze the psychometric validity of our measure of achievement goal orientation and learning strategy, we conducted an exploratory factor analysis with Varimax rotation and Kaiser normalization. The results of this assessment are shown in Table 2. Based on Eigenvalues > 1 , three distinct factors emerged. We had to exclude one item that represented “focus of attention” due to low factor loadings. After removing it, none of the items cross-loaded onto other factors (i.e., no crossloading of above .4). All other loadings remain stable and above .5. Thus, the three factors represent performance goal, mastery goal and deep learning strategy quite well and can be used for further analyses.

Next, we tested a model of bonus program-specific and student learning-related aspects of students’ future intention to participate in a bonus program again (see Figure 4). In addition to the three factors that stem out of the factor analysis (i.e. student learning-related aspects), we integrated two aspects that relate to the design of the bonus program as such, both measured with a single item based on discussions with students and tutors. We note that according to [BR07], the application of single item measures is appropriate, when the underlying aspects can be represented by a single item. The first item relates to

the difficulty of designing an exercise along with a solution (i.e. task difficulty) and reads: “I perceived developing exercises generally as difficult”. The second item pertains to fairness in the sense of a ratio between effort and outcome (i.e. adequate effort) and reads “The effort required to develop exercises stood in an adequate relation to the result in form of extra points”.

As the questionnaire started with standardized questions concerning the course as such, which were designed to align with the German system where ‘1’ indicates ‘very good’, we anchored the additional questions concerning the bonus program and learning strategies with ‘1’ = fully agree to ‘5’ = fully disagree to not bemuse participants. However, to better be able to interpret results in relation to the dependent variable future intention, which was coded ‘0’ = future intention NO and ‘1’ = future intention YES, we reverse coded all items prior to putting them into analyses. Based on this procedure, we observed that task difficulty ($M = 3.52$, $SD = 1.05$) was rated a little higher than adequate effort ($M = 2.82$, $SD = 1.21$), indicating that the tasks were generally perceived as demanding.

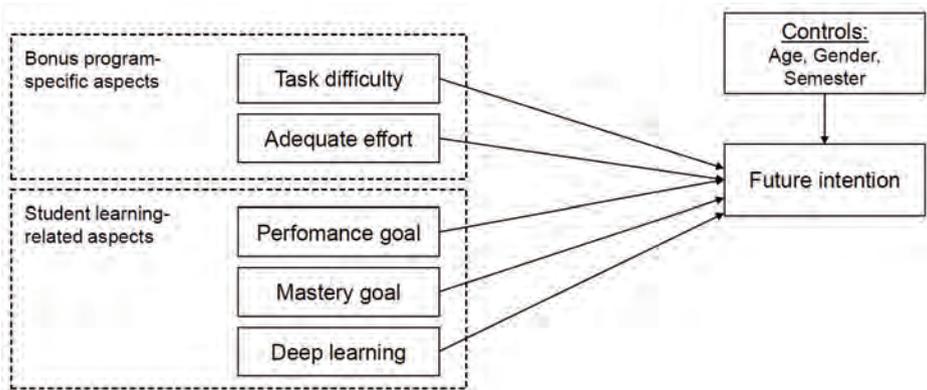


Fig. 2: Conceptual model

As our dependent variable is of a dichotomous nature, we applied logistic regressions to assess the effects of the antecedents of future intention to participate [HL04]. The logistic regression including all study participants ($n = 60$) expressed the predicted values as probabilities and the predicted proportion of students willing to participate in the future as the logistic model $\exp(X)/(1-\exp(X))$, where X is a linear function of the independent variables. As the results in Table 3 indicate, the overall model predicting future intention revealed good fit, as indicated by appropriate R^2 values. In addition, the [HL04] measure of overall fit was not significant ($\chi^2 = 5.412$; $df = 8$; $sig. = .713$). Finally, we can expect that 86.7% of all cases classified correctly, thus indicating an acceptable overall model fit. Table 3 provided the results for the test of factors affecting future intention to participate. None of the controls, that is, age, gender, and semester, showed a significant influence on future intention. In addition, neither performance goal orientation nor mastery goal orientation revealed any significant relationship with the dependent variable. Finally, of the two aspects reflecting bonus program-specific facets,

namely task difficulty and adequate effort, only adequate effort had a significant predictive power ($B = 3.03$, $\text{Exp}(B) = 20.586$); meaning that a one-unit increase in adequate effort increases the likelihood of future intention by 20.586.

	B	S.E.	Wald	Sig.	Exp(B)
<i>Co-variables</i>					
Age	.33	.37	.77	.380	1.385
Gender	-.83	1.21	.47	.493	.436
Semester	.38	.37	1.07	.300	1.465
<i>Predictor variables</i>					
Task difficulty	-.06	.49	.02	.895	.937
Adequate effort	3.03	1.22	6.17	.013	20.586
Performance goal	.26	.66	.16	.689	1.301
Mastery goal	.49	.60	.68	.410	1.637
Deep learning strategy	.52	.48	1.20	.274	1.684
Constant	-11.15	8.79	1.61	.205	.000
N	60				
Cox & Snell R ²	.319				
Nagelkerkes R ²	.558				
-2 log-Likelihood	287.714				

Tab. 3: Logistic regression results; DV: Future intention

5 Lessons learnt

The evaluation results show that students are satisfied with our approach of crowdsourcing-based learning on various dimensions, as indicated by a high number of students that are willing to participate in the bonus program again and an overall satisfaction level with course as such (including the impression of how well the lecturer performed) of 2.2 on a scale from 1 to 6. There are several lessons learnt on which we like to report. First, according to our first implementation of a crowdsourcing-based learning approach, bonus points are a sufficient incentive for students to participate in a crowdsourcing-based learning arrangement. Of course, setting this incentive too high will stimulate participation but not necessarily activate active learning among the entire group of learners to the same degree an “adequate” bonus would do. As adequate effort was rated comparatively moderate ($M = 2.82$, $SD = 1.21$, on a 5-point scale), we are confident to have balanced our demands and related bonus in a “fair” way. Nevertheless, future research could investigate how different incentives in form of bonus points affect satisfaction and learning success.

Second, the implementation of the approach was new to us. Therefore, we used comments of students after each round of exercises to continuously improve the approach. However, offering this opportunity when little teaching resources available is demanding, as the time between two waves of submitted bonus works is relatively short to implement changes to a satisfying degree.

Third, according to our impressions, students need as much documentation as possible. A detailed description of the task and an example exercise and solution for each task must be handed to the students. If students have to grade exercises (i.e., assigning a complexity score) then comparative example exercises with scores must be handed as well. Otherwise, students tend to submit “simple” exercises with high complexity scores and are disappointed when not receiving full points. Our evaluation schema benchmarked the design of the exercise (appropriateness of complexity score vs. complexity of the exercise) and the design of the solution (correctness, detailed documentation). Additional points were granted if the submitted exercises differed from the example exercise pointing to creativity of the student.

Finally, within several discussions with experts and administration staff, we learnt that our crowdsourcing-based learning approach is particularly useful when exams are only marked with “pass” and “failed” and not with grades like 1.0, 1.3, 1.7, 2.0, 2.3, 2.7, ... 5.0. The assessment schema “pass” or “failed” is currently superior as it does more comply with current German exam requirements (e.g., requirement that each student must receive an exam that equals other exams in terms of complexity).

6 Conclusion and Future Directions

This paper presented a crowdsourcing-based learning approach for active learning including an evaluation study. The approach has been illustrated on the process of exam preparation. Students were asked along the complete semester to design exam exercises and solutions for selected topics of the lecture “Information Systems”. The self-creation of exam exercises is the foundation of a crowdsourcing-based learning approach. Our empirical evaluation has shown that an exam bonus is a sufficient incentive. However, the effort required to develop exercises should be in an adequate relationship to the result in form of bonus points. However, several technical, conceptual (exam design), and administrative challenges must be solved to unravel the full potential of the approach.

The first challenge to be addressed is the comparability of the automatically generated exam questions. The current course content management system allows to generate an exam from predefined questions or to compose the exam from randomly selected questions but is limited when it comes to assigning complexity scores. To make the (automatically composed) exams comparable, the self-assessed complexity of exercises by students can be taken into consideration as well as statistical measures, which rate the fulfillment of learning objectives per question type. According to [SG06] students are able to grade in a similar way as teachers when the scoring rubric is trained to them. In our context, a star rating system with a scale from 1 (= low) to 5 (=high) seems to be an

appropriate grading system since the use of the rating system is known by students (and also no training effort would be necessary). Thus, first the student self-assesses his/her exercise and then peers assess the complexity of the exercise resulting in an average star value, which gives a rough picture of the complexity of the exercise. The average star value is considered in the composition of exams.

The second challenge addresses the effort for *quality assurance* (revision) of questions by the lecturer. This effort should be reasonable (lower) compared to the effort for marking of paper-written tests. Students already reduce the revision effort of the lecturer when revising the exercises of their peers (they improve the language as well as the concepts). The effort decreases from one student to the next. The final quality assurance, however, should be a task of the lecturer. A future direction would be to investigate influences of the review process on quality of the revision of exercises. In addition, research could ask if there is a correlation between the number of revision cycles and the knowledge and expertise of students.

The third challenge of the crowdsourcing-based learning approach is the compatibility between automatically generated exams and University's examination regulations. It is likely that changes to examination regulations are necessary before individual exams can be offered to students. These changes affect Intellectual Property (IP) (i.e. who is the user of the exercises and who the owner?). Students create an exercise and the solution. Then peers revise the exercises and solutions. Finally, the lecturer approves the quality (this might also lead to adaptations of question/answer pairs). Consequently, a discussion might arise that the final exercise (which was iteratively revised) belongs to several parties (the student, peers and lecturer) and how these groups enforce their IP rights. Similar discussions arose around "owners of peer reviews" ([Be10, Cr15]) and studies have been conducted to find answers. Comparative analysis investigating correlations between student exercise quality and the final review quality are still outstanding, but would be fundamental to a crowdsourcing-based learning approach. Thus, we call for future research that combines crowdsourcing-based active learning with IP issues.

Bibliography

- [AA88] Ames, C.; Archer, J.: Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology* 80/3, S. 260–267, 1988.
- [An11] Anderson, M.: Crowdsourcing Higher Education: A Design Proposal for Distributed Learning. *Journal of Online Learning and Teaching*, 7/4, S. 576–590, 2011.
- [Be10] Bernstein, D.: Peer Review and Evaluation of the Intellectual Work of Teaching. *Change. the Magazine of Higher Education* 40/2. S. 48–51, 2010.
- [Bo01] Boud, D.: Introduction: Making the Move to Peer Learning. in: *Peer learning in higher education*, Kogan Page, S. 1–14, 2001.
- [BR07] Bergkvist, L.; Rossiter, J.R.: The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research*, 44/2, S. 175–184,

2007.

- [Ce16] Center of Research and Learning: Active Learning. University of Michigan, Center for Research on Learning and Teaching, 2016, <http://www.crlt.umich.edu/tstrategies/tsal>, Stand: 18.07.2017.
- [CM12] Corneli, J.; Mikroyannidis, A.: Crowdsourcing education on the Web. A rolebased analysis of online learning communities.” in: Collaborative Learning 2.0: Open Educational Resources, IGI Global, S. 272–286, 2012.
- [Cr15] Crotti, D.: The Problem(s) With Credit for Peer Review, 2015, <https://scholarlykitchen.sspnet.org/2015/06/17/the-problems-with-credit-for-peer-review/>, Stand: 18.07.2017.
- [De09] Dermo, J.: e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology* 40/2, S. 203–214, 2009.
- [HL04] Hosmer Jr, D. W.; Lemeshow, S: Applied logistic regression. John Wiley&Sons, 2004.
- [Ho06] Howe, J.: The rise of crowdsourcing. *Wired Magazine* 14/6, 2006.
- [KB16] Koschmider, A.; Buschfeld, D: Shifting the process of exam preparation towards active learning: A crowdsourcing based approach. GI-Jahrestagung 2016, S. 1017–1022, 2016.
- [Le09] Leimeister, J. M.; Huber, M.; Bretschneider, U.; Krömer, H.: Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. *Journal of Management Information Systems*, 26 /1, S. 197–224, 2009.
- [LL15] Lehmann, K.; Leimeister, J.M.: Theory-Driven Design of an IT-Based Peer Assessment to Assess High Cognitive Levels of Educational Objectives in Large-Scale Learning Services. In: ECIS 2015, Münster, 2015.
- [LN05] Lomas, L.; Nicholls, G.: Peer Review and Evaluation of the Intellectual Work of Teaching. *Quality in Higher Education* 11/2, S. 137–149, 2005.
- [LN11] Lai, Y. C.; Ng, E.: Using wikis to develop student teachers' learning, teaching, and assessment capabilities. *The Internet and Higher Education* 14/1, S. 15–26, 2011.
- [No10] Novak, J. D.: Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations. Routledge, New York, 2010.
- [Sc15] Schlagwein, D.: Students as Reviewers and Lecturers as Editors: The Peer Review with Scaffolded Assignments Model. ICIS 2015, AISNet, Fort Worth, 2015.
- [SG06] Sadler, P.M.; Good, E.: The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment* 11/1, S. 1–31, 2006.
- [We12] Weld, D. S.; Adar, E.; Chilton, L.; Hoffmann, R.; Horvitz, E.; Koch, M.; Landay, J.; Mausam, C.H.L.: Personalized Online Education - A Crowdsourcing Challenge, 2012.
- [Wi16] Wibowo, S. et al.: A Pilot Study of an Electronic Exam System at an Australian University. *Journal of Educational Technology Systems* 45/1, S. 5–33, 2016.

Akzeptanz von e-Learning-Strukturen in Hygieneweiterbildungen

Michael Meng¹, Thorsten Hey², Daniel Peter³, Frauke Mattner⁴, Christoph Igel⁵ und Christiane Kugler⁶

Abstract: Das Nutzungsverhalten von Teilnehmenden eines Web 2.0 basierten privat social Network e-Learning Systems, im Kontext einer Weiterbildung von Hygienefachkräften zur Verbesserung der Ausbildungssituation von hygienebeauftragten Pflegenden, wurde deskriptiv und explorativ untersucht. N=126 Hygienefachkräfte nahmen an einer Evaluationsstudie im Zeitraum von sieben Monaten teil. Aufgrund einer zeitversetzten Etablierung der e-Learning-Struktur, wurden N=76 Nutzer zur Nutzung freigeschaltet. Die Teilnehmenden hatten eine signifikant größere computerbezogene Selbstwirksamkeitserwartung als die literaturbasierte Normstichprobe ($t(66)=13.24$, $p=.000$). Weder in Bezug auf Alter ($u=285.50$, $z=-1.95$, $p=.051$) noch Geschlecht ($u=243.50$, $z=-0.32$, $p=.75$) konnten signifikante Unterschiede in der computerbezogenen Selbstwirksamkeitserwartung nachgewiesen werden. Die Nutzer haben insgesamt 5656 Minuten im System verbracht und 9542 Aktivitäten durchgeführt. Aufgrund der Datenlage kann der vorsichtige Schluss gezogen werden, dass die Web 2.0 basierte e-Learning-Struktur von Gesundheitsakteuren im Bereich der Krankenhaushygiene im Kontext eines Weiterbildungscurriculums angenommen wurde und für eine künftige Nutzung als Ressource zur Verfügung steht.

Keywords: Web 2.0, Hygiene, Curriculum, e-Learning, computerbezogene Selbstwirksamkeitserwartung

1 Hintergrund

Schätzungsweise 400.000 – 600.000 Patienten in Deutschland erleiden jährlich eine nosokomiale, im Krankenhaus erworbene, Infektion, wobei etwa 10.000 – 15.000 Patienten daran versterben [HLS15]. Zur Verbesserung der Ausbildungssituation von hygie-

¹ Universität Witten/Herdecke, Department für Pflegewissenschaft, Stockumer Straße 10, 58453 Witten, michael.meng@uni-wh.de

² Deutsches Forschungszentrum für Künstliche Intelligenz, Educational Technology Lab, Alt-Moabit 91c, 10559 Berlin, thorsten.hey@dfki.de

³ Klinken der Stadt Köln gGmbH, Institut für Hygiene, Ostmerheimer Str. 200, 51109 Köln, peterd@kliniken-koeln.de

⁴ Klinken der Stadt Köln gGmbH, Institut für Hygiene, Ostmerheimer Str. 200, 51109 Köln, mattnerf@kliniken-koeln.de

⁵ Deutsches Forschungszentrum für Künstliche Intelligenz, Educational Technology Lab, Alt-Moabit 91c, 10559 Berlin, christoph.igel@dfki.de

⁶ Universität Freiburg, Medizinische Fakultät, Fehrenbachallee 8, 79085 Freiburg, christiane.kugler@uniklinik-freiburg.de

Dieser Beitrag entstand im Rahmen eines durch das Bundesministerium für Gesundheit geförderten Projektes (IIA5-2512FSB102).

nebeauftragten Pflegenden, den Hygieneakteuren in der pflegerischen Versorgung, haben wir ein Curriculum zur Weiterbildung von Hygienefachkräften entwickelt [Kr17]. Begleitend zu den Präsenzphasen der Weiterbildung wurde ein Web 2.0 basiertes privat social Network (PSN) als e-Learning-Struktur eingesetzt. Ein PSN ist ein virtuelles soziales Netz zum digitalen Informationsaustausch. Realisiert wurde dies durch Elgg, einer frei zugänglichen social Networking Software. Mit Elgg wurde ein PSN erstellt. Für jede Gruppe der Präsenzphasen wurde eine Gruppe im PSN angelegt. Die Teilnehmenden konnten innerhalb ihrer Gruppe kommunizieren und Daten und Wissen austauschen. Bestandteil der Entwicklung und Pilottesting der Weiterbildung ist eine Evaluation des Curriculums, die an das evaluative Framework von Kirkpatrick und Kirkpatrick, dem 4-Ebenen Model und dem New World Model, ausgerichtet wurde und auf mehreren Ebenen durchgeführt wird [KK06; KK16]. Der traditionelle Begriff des e-Learning, *e-Learning 1.0*, kann nach Ebner [Eb07] als eine Erweiterung des klassischen Unterrichts verstanden werden. Durch die fortlaufende technische Entwicklung bietet das *Web 2.0* nun eine Erweiterung um eine soziale Komponente, welche in Bezug auf Lernprozesse von großer Bedeutung ist [GZ08]. Das *Web 2.0* ist nach Hussain [Hu12] eine dynamischere Version des *Web 1.0*. Das *e-Learning 2.0* resultiert daraus als ein kollaborativer Weg des Lernens, wobei Kommunikation multidirektional erfolgt und Wissensinhalte gemeinsam im sozialen Kontext konstruiert werden [Hu12]. Cassidy und Eachus (2002) [CE02] verdeutlichen, dass hierfür die computerbezogene Selbstwirksamkeitserwartung als ein wichtiger Faktor zur Akzeptanz und einer effektiven Nutzung neuer e-Learning-Strukturen zu berücksichtigen ist. Im Bereich von Pflegeschulungen werden in den letzten Jahren zunehmend e-Learning-Strukturen eingesetzt [III14]. Ein systematischer Review von Mccutcheon et al. (2015) zeigte, dass e-Learning in der Ausbildung von Pflegenden, im Vergleich zu traditionellen Lehrmethoden, vergleichbare Ergebnisse, u.a. in Bezug auf den Erwerb von Wissen und Fertigkeiten, liefert [Mc15]. Gleichzeitig wird gefordert, dass e-Learning als Web 2.0 basierte Struktur in Weiterbildungen in der Pflege eingesetzt werden sollte [La13]. Der vorliegende Beitrag hat zum Ziel, das Nutzungsverhalten von Teilnehmenden eines Web 2.0 basierten PSN e-Learning Systems, im Kontext einer Weiterbildung von Hygienefachkräften zur Verbesserung der Ausbildungssituation von hygienebeauftragten Pflegenden, deskriptiv und explorativ zu untersuchen.

2 Methode

2.1 Design

In einem mehrstufigen Studiendesign wurden Kenngrößen zur quantitativen und qualitativen Evaluation des Curriculums analysiert. Die Studie wurde von einer multiprofessionellen Arbeitsgruppe durchgeführt (*Medizin, Pflegewissenschaft, Psychologie*). Zur Durchführung der Studie liegt ein positiver Bescheid durch die zuständige Ethikkommission vor.

2.2 Fragebogen zur computerbezogenen Selbstwirksamkeitserwartung

Zur Erfassung einer interpretatorischen Baseline zur Analyse der Daten und zur deskriptiven Beschreibung der gegebenen Stichprobe, wurde ein Fragebogen zur Messung der computerbezogenen Selbstwirksamkeitserwartung der Teilnehmenden mit 30 Items eingesetzt [CE02]. In der vorliegenden Arbeit wurde die validierte und ins Deutsche übersetzte Version des Fragebogens zur Erhebung der Daten verwendet, wobei hohe Werte hohe Selbstwirksamkeitserwartung bedeuten [SB09].

2.3 Deskriptive und explorative Analyse der Nutzerdaten anhand einer Datenbankabbilddatei

Die Analyse des Nutzungsverhaltens der Teilnehmenden in der implementierten Web 2.0 basierten e-Learning-Struktur wurde anhand einer auf dem Server liegenden Datenbankabbilddatei durchgeführt. Dabei wurden mehrere Variablen zur deskriptiven und explorativen Analyse der Nutzung erzeugt. Die abgedeckten Kategorien der Nutzung des Systems waren auf Nutzer-Ebene die *Gesamtzeit der Nutzung des Systems*, die *Gesamtanzahl von Aktivitäten im System* und die Anzahl der Aktivitäten *Login*, *Logout*, *Create Annotate*, *Update* und *Delete*. Zur Übersicht wurden Summenwerte und Mittelwerte der Kategorien errechnet.

2.4 Datensammlung

Das neue Weiterbildungsangebot richtete sich an Hygienefachkräfte. Die Datensammlung für das gesamte zu evaluierende Projekt erfolgte von 2013 bis 2016. In diesem Beitrag werden Daten im Zeitraum der ersten sieben Monate nach Inbetriebnahme betrachtet. Das Curriculum ist modular mit vier ganztägigen Präsenzphasen aufgebaut. Zwischen den ersten drei Modultagen verging jeweils ein Monat, zwischen dem dritten und vierten Modultag mindestens drei Monate. Ab dem ersten Präsenztag wurden die Teilnehmenden in einem Web 2.0 basierten PSN registriert und konnten innerhalb ihrer Weiterbildungskohorte interagieren. Die Sammlung der Nutzerdaten erfolgt über ein Datenbankabbild des Servers.

2.5 Datenauswertung

Die deskriptive und explorative Datenanalyse der Nutzerdaten des Zeitraumes von sieben Monaten erfolgte über die Datenbankabbilddatei des Servers. Die Datenbankabbilddatei wurde hierbei über ein zusätzliches, eigens programmiertes Programm ausgelesen, womit die Daten kumuliert angezeigt werden können. Die inferenzstatistischen Auswertungen erfolgten mithilfe des Statistikprogrammes IBM SPSS 22. Fehlende Werte einzelner Variablen werden ausgewiesen. Das Signifikanzniveau wurde a priori auf $p < 0.05$ festgelegt.

3 Ergebnisse

3.1 Stichprobe

Insgesamt haben 126 Hygienefachkräfte aus unterschiedlichen Krankenhäusern bundesweit an der Studie teilgenommen. Die formale Zugangsvoraussetzung zur Teilnahme an der Evaluationsstudie war eine abgeschlossene Berufsausbildung in der Gesundheits- und Krankenpflege und eine anschließende, erfolgreich absolvierte Weiterbildung als Hygienefachkraft. 85.6% der Teilnehmenden waren weiblich, 14.4% waren männlich. 1.6% der Teilnehmenden waren unter 30 Jahre alt, 16% waren zwischen 31 und 40 Jahre alt, 47.2% zwischen 41 und 50 Jahre und 35.2% waren über 50 Jahre alt. Diese Altersverteilung und Geschlechtsverteilung beschreibt eine repräsentative Stichprobe in Bezug auf Pflegende in Deutschland [St13]. Die meisten Teilnehmenden waren hochqualifizierte Fachkräfte, mit über 10 Jahren Berufserfahrung (94.4%). 3.2% hatten 6-10 Jahre Berufserfahrung und 1.6% 5-6 Jahre Berufserfahrung. Die zur deskriptiven und explorativen Datenanalyse der Nutzerdaten genutzte Datenbankabbilddatei, welche einen Zeitraum von sieben Monaten abbildet, enthält 76 Nutzer. Die abweichende Nutzerzahl des e-Learning Systems von N=76 Nutzer zu der untersuchten Grundgesamtheit von N=126 Hygienefachkräften ist auf die entwicklungsbedingte, zeitversetzte Etablierung der Web 2.0 basierten e-Learning-Struktur zum Beginn der Präsenzphasen zurückzuführen. Um eine standardisierte Testsituation zu gewährleisten, wurde darauf verzichtet, die bereits im Weiterbildungsprozess befindlichen Weiterbildungsläufe in die e-Learning-Struktur nachträglich aufzunehmen.

3.2 Computerbezogene Selbstwirksamkeitserwartung

Insgesamt füllten N=67 Teilnehmende den Computer User Self-Efficacy Scale Fragebogen in deutscher Übersetzung (CUSE-D) [SB09] aus. Dies entspricht einer Rücklaufquote von 88.2%. Die Teilnehmenden erreichten im arithmetischen Mittel (MW) einen Wert von $MW=135.22$ ($SD=20.54$) in Bezug auf die selbst eingeschätzte computerbezogene Selbstwirksamkeitserwartung. Die zugehörige Normstichprobe von Pflegenden hat einen entsprechenden Mittelwert von $MW=101.52$ ($SD=30.50$) [CE02]. Zur Überprüfung eines statistisch signifikanten Unterschieds der Mittelwerte wurde ein T-Test für eine Stichprobe gerechnet. Es zeigt sich ein signifikanter Unterschied zwischen dem Mittelwert der erhobenen versus literaturbasierten Stichprobe ($t(66)=13.24$, $p= .000$) zugunsten der erhobenen Stichprobe. Das Alter ($n=66$) der Teilnehmenden wurde zur weiteren Analyse entlang des Medians $m=47$ dichotomisiert. $n=29$ Teilnehmende waren jünger als 47 Jahre, $n=37$ waren 47 Jahre oder älter. Ein U-Test nach Mann-Whitney ergab keinen signifikanten Unterschied der dichotomisierten Altersgruppen ($u=285.50$, $z= -1.95$, $p= .051$). Zuletzt wurde ein U-Test zum Vergleich der Geschlechter (N=67) gerechnet. Hier zeigte sich kein signifikanter Unterschied zwischen Frauen ($n=58$) und Männern ($n=9$) ($u=243.50$, $z= -0.32$, $p= .75$).

3.3 Deskriptive und explorative Beschreibung des Nutzungsverhaltens

Die N=76 Nutzer zeigten, im Zeitraum von sieben Monaten, eine gesamte Online-Zeit von 5656 Minuten (94.27 Stunden). Im Mittelwert entspricht dies einer durchschnittlichen Nutzungszeit von $MW=74.4$ Minuten pro Person über die Beobachtungszeit. Die Teilnehmenden haben insgesamt 9542 Aktivitäten in dem e-Learning System durchgeführt, was im Mittel $MW=125.6$ Aktivitäten pro Person bedeutet. Die 9542 Aktivitäten lassen sich in insgesamt 383 Aktivitäten *Login* ($MW=5.0$ *Login* pro Person), 291 Aktivitäten *Logout* ($MW=3.8$ *Logout* pro Person), 8473 Aktivitäten *Create* ($MW=111.5$ *Create* pro Person), 54 Aktivitäten *Annotate* ($MW=0.7$ *Annotate* pro Person), 156 Aktivitäten *Update* ($MW=2.1$ *Update* pro Person) und 181 Aktivitäten *Delete* ($MW=2.4$ *Delete* pro Person) differenzieren.

4 Diskussion

Insgesamt nahmen N=126 Hygienefachkräfte an der Evaluationsstudie teil, wobei N=76 Nutzer des e-Learning Systems insgesamt 5656 Minuten im System verbracht und 9542 Aktivitäten in dem Zeitraum von sieben Monaten durchgeführt haben. Die Analyse von Nutzungsdaten über Datenbankabbilddateien wird in der Wissenschaft, neben der kommerziellen Verwendung, angewendet [We14]. Die Analyse des Nutzungsverhaltens konnte zeigen, dass die Teilnehmenden aktiv mit dem System interagiert haben. Im Kontext der relativen zeitlichen Verortung der Datenbankabbilddatei zum Projektstart und des relativ kurzen Zeitraumes von sieben Monaten, kann der vorsichtige Schluss gezogen werden, dass eine gewisse Akzeptanz der Web 2.0 basierten e-Learning-Struktur im Kontext der Weiterbildung stattgefunden hat. Diese These kann durch die Betrachtung der computerbezogenen Selbstwirksamkeitserwartung gestützt werden. Durch die, im Vergleich zur Literatur, signifikant erhöhte computerbezogene Selbstwirksamkeitserwartung, kann der vorsichtige Schluss gezogen werden, dass die vorliegende Stichprobe wenige hemmende Faktoren bei der Nutzung einer Web 2.0 basierten e-Learning-Struktur erlebt. Die Unterschiede in der Login- und Logout- Häufigkeit, lassen den vorsichtigen Schluss zu, dass eine zusätzliche Aufklärung im Bereich der Sicherheit im Internet überdacht werden sollte. Grenzen solcher Datenbankabbild-Analysemethoden liegen vor allem in der Art der Rohdaten. Aussagen über den inhaltlichen Gegenstand einer definierten Aktionen, beispielsweise *create*, können nur begrenzt getroffen werden. Daher sollte diesem Gesichtspunkt bei der Interpretation von Nutzungsdaten besonders, zum Beispiel durch Methodenkombinationen, Rechnung getragen werden [We14]. Eine Limitation des vorliegenden Beitrages ist die kurze Zeitspanne, welche die Datenbankabbilddatei umfasst. In folgenden Untersuchungen sollte ein längerer Zeitraum zur Analyse gewählt werden, um den Teilnehmenden mehr Aktivität im e-Learning zu ermöglichen. Zudem sollte in folgenden Untersuchungen eine größere Kohorte untersucht werden. Beide Faktoren haben potentiell einen Einfluss auf die aktuell relativ geringen durchschnittlichen Nutzungsdauern und die Möglichkeiten zur Analyse und Interpretation des Nutzungsverhaltens.

Literaturverzeichnis

- [CE02] Cassidy, S.; Eachus, P.: Developing the Computer User Self-Efficacy (CUSE) Scale. Investigating the Relationship between Computer Self-Efficacy, Gender and Experience with Computers. *Journal of Educational Computing Research* 26/2, S. 133-153, 2002.
- [Eb07] Ebner, M.: E-Learning 2.0 = e-Learning 1.0 + Web 2.0? In *Second International Conference on Availability, Reliability and Security (ARES'07)*, S. 1-5, 2007.
- [GZ08] Gerrig, R.; Zimbardo, P.: *Psychologie*. 18. neubearbeitete Auflage, Pearson Studium, München, 2008.
- [HLS15] Häfner, H.; Lemmen, S.; Scheithauer, S.: Nosokomiale Infektionen – Epidemiologie, Diagnostik, Therapie und Prävention. In (Marx, G.; Muhl, E.; Zacharowski, K.; Zeuzem, S., Hrsg.): *Die Intensivmedizin*. Springer, Berlin, S. 949-970, 2015.
- [Hu12] Hussain, F.: E-LEARNING 3.0 = E-LEARNING 2.0 + WEB 3.0? In *IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2012)*, S. 11-18, 2012.
- [Ill4] Ilott, I.; Bennett, B.; Gerrish, K.; Pownall, S.; Jones, A.; Garth, A.: Evaluating a novel approach to enhancing dysphagia management: workplace-based, blended e-learning. *Journal of Clinical Nursing* 23/9-10, S. 1354-1364, 2014.
- [KK06] Kirkpatrick, D.L.; Kirkpatrick, J.D.: *Evaluating Training Programs: The Four Levels*. 3. Auflage, Berrett-Koehler Publishers, San Francisco, 2006.
- [KK16] Kirkpatrick, J.D.; Kirkpatrick, W.K.: *Four Levels of Training Evaluation*. ATD Press., Alexandria, 2016.
- [Kr17] Krüger, C.; Meng, M.; Mattner, F.; Peter, D.; Kugler C.: Entwicklung eines Curriculums für Hygienefachkräfte zur Qualifikation von hygienebeauftragten Pflegenden - Ergebnisse von Fokusgruppeninterviews mit Stationsleitungen und Hygienefachkräften in deutschen Akutkrankenhäusern. *Pflege - Die wissenschaftliche Zeitschrift für Pflegeberufe* 30 (published online), S. 1-11, 2017.
- [La13] Labeau, S.O.: Is there a place for e-learning in infection prevention? *Australian Critical Care* 26/4, S. 167-172, 2013.
- [Mc15] McCutcheon, K.; Lohan, M.; Traynor, M.; Martin, D.: A systematic review evaluating the impact of online or blended learning vs. face-to-face learning of clinical skills in undergraduate nurse education. *Journal of Advanced Nursing* 71/2, S. 255-270, 2015.
- [SB09] Spannagel, C.; Bescherer, C.: Computerbezogene Selbstwirksamkeitserwartung in Lehrveranstaltungen mit Computernutzung. *Notes on Educational Informatics - Section A: Concepts and Techniques* 5/1, S. 23-43, 2009.
- [St13] Statistisches Bundesamt (DESTATIS): *Health, personnel*. Fachserie 12 Reihe 7.3.1. Statistisches Bundesamt, Wiesbaden, 2013.
- [We14] Welker, M.; Taddicken, M.; Schmidt, JH; Jakob, N.: *Handbuch Online-Forschung*, Köln, 2014.

Zielgruppengerechte App-Icons für Seniorinnen und Senioren

Svenja Noichl¹, Nadine Bergner¹ und Ulrik Schroeder¹

Abstract: Immer mehr Seniorinnen und Senioren nutzen Smartphones und Tablets. Lern-Apps können Seniorinnen und Senioren beim Erlernen der Nutzung dieser Technologien unterstützen. Wichtig bei der Nutzung von Apps ist das Verständnis der verwendeten Icons. Diese sollten intuitiv und unmissverständlich sein. Bei unerfahrenen Nutzerinnen und Nutzern, insbesondere bei Seniorinnen und Senioren ist dies jedoch nicht immer gegeben. In einer ersten Stufe soll ein Icon-Set entwickelt werden, welches auch von der Zielgruppe der Seniorinnen und Senioren intuitiv verstanden wird und in zukünftig zu entwickelnden Lern-Apps eingesetzt werden soll. In einer Vorstudie wurde eine heterogene Gruppe zur Verständlichkeit ausgewählter Icons aus dem Material Icon Set befragt. Die Befragung ergab, dass es zum einen Icons gibt, die verstanden werden, und zum anderen Icons, wie z. B. das empfohlene Icon für „eingeschalteter Wecker“ missinterpretiert werden.

Keywords: Seniorinnen und Senioren, Verständnis von Icons, Smartphone, Tablet, Vorstudie, App-Design, Nutzerinterface

1 Einleitung & Motivation

Immer mehr Seniorinnen und Senioren nutzen Smartphones und Tablets. Nutzten 2014 noch ungefähr 32% der Deutschen im Alter zwischen 50 und 59 Jahren ein Smartphone, hat sich dieser Anteil bis 2016 fast verdoppelt (etwa 60%). Bei Personen zwischen 60 und 69 sowie bei den ab 70-jährigen hat sich der Anteil sogar fast verdreifacht (60-69 Jahre: von 13% auf 37%; 70+ Jahre: von 4% auf 14%) [GfI6]. Zu beachten ist an dieser Stelle, dass in dieser Zielgruppe nicht nur das Alter selbst, sondern auch die individuellen Hintergründe eine entscheidende Rolle spielen. Bei der Nutzung von Apps durch Seniorinnen und Senioren ergeben sich häufig Probleme, da die meisten Apps für jüngere Menschen entwickelt werden, die häufig bereits über Vorerfahrungen verfügen. Im Rahmen der Forschung sollen didaktische und technische Konzeptionen von E-Learning Maßnahmen für Seniorinnen und Senioren entwickelt werden. Diese sollen in kleinen Lern-Apps umgesetzt werden, damit auch diese Zielgruppe die Vorteile von digitalen Technologien nutzen und beispielsweise durch E-Learning-Elemente ihre geistige und körperliche Fitness stärken kann. Um eine Interaktion mit den Geräten zu ermöglichen, ist die richtige Interpretation der verwendeten Icons als Kennzeichnung für ausführbare Aktionen zwingend notwendig.

¹ RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, 52074 Aachen {noichl, bergner, schroeder}@informatik.rwth-aachen.de

Dieses Paper beschreibt eine Vorstudie zur Verständlichkeit der weit verbreiteten Google Material Icons mit einer heterogenen Gruppe von Tablet Einsteigerinnen und Einsteigern. In Kapitel 2 werden bisherige Untersuchungen in diesem Kontext, sowie die Verbreitung von Icons betrachtet. Kapitel 3 beschäftigt sich mit Design, Durchführung und Ergebnissen der Vorstudie. Basierend auf den Ergebnissen soll in weiteren Studien, in engem Kontakt mit der Zielgruppe, ein Icon Set entstehen, welches für die Zielgruppe der Seniorinnen und Senioren angepasst ist.

2 Related Work

Mobile Geräte bieten ein großes Potential in der Unterstützung von älteren Menschen, beispielsweise bei der Erinnerung an die Medikamenteneinnahme oder dem Kontakt zu Angehörigen. Jedoch haben sie häufig Probleme, die verwendeten Icons zu interpretieren.

Eine Studie in Kanada ergab 2011, dass Beschriftungen beim Verständnis helfen können. Die Studie wurde durchgeführt, bevor mit dem Google Material Design und den dazu gehörenden Icons ein einheitliches Schema zur Gestaltung von Apps bereitgestellt wurde. Weiter wird diskutiert, wie die Ergebnisse Designern solcher Icons helfen können, für ältere Menschen verständlichere Icons zu entwerfen. Es scheint sinnvoll, die Nutzer aus einem vorgegebenen Angebot selbst wählen zu lassen, welches Icon für eine bestimmte Funktionalität angezeigt werden soll. Denn jeder kann ein anderes Symbol z. B. mit der Funktion „Einstellungen“ assoziieren. Auf diese Weise ist eine Erleichterung im Umgang mit Icons möglich. [LMG11]

Aktuelle Untersuchungen in dem österreichischen Forschungsprojekt *mobi.senior.A* zeigen, dass unerfahrene Nutzer die Bedeutung der in der Praxis gängigen Icons nicht intuitiv erkennen und verstehen. Dies trifft insbesondere auf die Zielgruppe der Seniorinnen und Senioren zu. Unverständlichkeit der Icons führt dabei zu schlechterer Bedienfähigkeit. Aus diesen Gründen wird zu einer konsistenten, appübergreifenden Nutzung von Icons für die wichtigsten Funktionalitäten geraten. [Er14, EX16]

Das Ziel, plattformübergreifend einheitliche Icons in Apps zu bringen, verfolgt Google mit den Material Icons, welche Teil des Material Design Konzepts sind. Ziel dieses Konzepts ist "create a visual language that synthesizes classic principles of good design with the innovation and possibility of technology and science" [Ma17]. Google stellt über 900 Icons zur Verfügung. Das Design dieser Icons soll „readability and clarity even at small sizes“ [Ma17] sicherstellen.

3 Verständlichkeitstest mit Expertengruppe

Für die Vorstudie wurden Icons aus dem Set der Material Icons verwendet. In den 50

meistgenutzten Android-Apps Deutschlands aus dem Jahr 2016 nutzen 25 Apps diese Icons. Darunter sind 8 Apps von Google selbst, allein 3 Apps befinden sich in den Top 5 der meistgenutzten Android-Apps. [ME16]

Auch bei den sieben im Online Ratgeber basenio empfohlenen Apps für Senioren (BIG Launcher, WhatsApp, Skype, Talk, Lumosity, MediSafe, MySugr) sind die Material Icons weit verbreitet. Fünf der Apps nutzen sie. [Ba16]

Die Vorstudie wurde als Teil eines Tabletworkshops durchgeführt, in welchem, neben der Befragung zur Verständlichkeit von Icons, kleine Lern-Apps eingesetzt wurden, um den Teilnehmerinnen und Teilnehmern Gesten, Eingabemöglichkeiten sowie die Installation einer App direkt am Gerät näher zu bringen. An dem Workshop nahmen sieben Personen im Alter zwischen 40 und 81 Jahren teil (<60 Jahre: 2 Personen; >70 Jahre: 5 Personen). Darunter waren vier Frauen und drei Männer. Drei der Teilnehmerinnen und Teilnehmer hatten keinerlei Vorerfahrung mit Smartphones oder Tablets, zwei Personen gaben an gelegentlich ein Android-Gerät zu nutzen, eine Person nutzte regelmäßig ein Android-Smartphone und eine weitere ein iPhone.

Das Ziel der Vorstudie bestand darin, die Verständlichkeit der gängigen Icons zu prüfen und herauszufinden, bei welchen Icons kein intuitives Verständnis gegeben ist. Die Forschungsfragen, welche durch die Vorstudie und insbesondere den Folgestudien beantwortet werden sollen, lauten:

1. Welche Icons werden von unerfahrenen Nutzerinnen und Nutzern und insbesondere von Seniorinnen und Senioren nicht intuitiv verstanden?
2. Für welche Bedeutungen ist es sinnvoll, Nutzerinnen und Nutzer individuell das zu verwendende Icon aus einer vorgegebenen Auswahl bestimmen zu lassen, da sie individuell unterschiedliche Icons damit assoziieren?
3. Wie können bestehende Icons verändert werden, damit Seniorinnen und Senioren diese besser verstehen?

Den Teilnehmerinnen und Teilnehmern wurden insgesamt drei Aufgaben zur Verständlichkeit der Icons gestellt. Bei den Aufgaben 1 und 2 waren 48 Icons aus dem Google Material Icon-Set gegeben, welche jeweils zugeordnet wurden. Die Icons lagen dabei in Papierform vor und waren mit einer Nummer versehen, um die Verständigung in der Gruppe zu vereinfachen. Zusätzlich wurden alle 48 Icons während der gesamten Durchführung an die Wand projiziert.

Nach einer individuellen Betrachtung der Icons wurden die Teilnehmerinnen und Teilnehmer in einer offenen Diskussionsrunde gebeten diejenigen Icons zu nennen, welche ihrer Meinung nach nicht unmittelbar verständlich sind. Am Ende der Diskussion wurden 14 Icons (Abbildung 1) genannt. Die Icons 28 und 32 erschienen ausschließlich den Gelegenheitsnutzern und Neulingen unverständlich. Den erfahrenen Nutzern waren diese Icons bekannt und sie nannten deren Bedeutung im Zusammenhang mit den von ihnen genutzten Apps. Icon 11 wurde mit „Was soll das denn sein? Vielleicht ein

Trockner?“ kommentiert und zu Icon 40 wurde bemerkt „keine Ahnung, warum da ein Z drin ist“. Diese erste Aufgabe zeigte, dass etwa ein Drittel der vorgelegten Icons laut Meinung der Befragten besser nicht benutzt werden sollte, da sie nicht unmittelbar deren Bedeutung erkennen können.

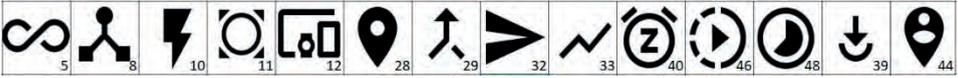


Abb. 1: Icons, die laut Expertengruppe nicht eingesetzt werden sollten

Zur Beantwortung der Fragestellung „Welches Icon gehört zur gegebenen Bedeutung?“, erhielten die Teilnehmerinnen und Teilnehmer ein Arbeitsblatt, auf welchem die fünf Bedeutungen „Person zum Adressbuch hinzufügen“, „Markierung für einen bestimmten Ort auf einer Karte“, „Navigation zu einem Ort“, „Eingeschalteter Wecker/Alarm“, „Bildschirm/Tablet drehen“ vorgegeben waren. Zu jeder Bedeutung gab es zwei Felder, zum einen für das vermutete Icon und zum anderen für das gewünschte Icon. Diese Tabelle wurde zunächst individuell von den Teilnehmerinnen und Teilnehmern mit den Icons gefüllt, Mehrfachnennungen waren dabei erlaubt. Im Anschluss folgte erneut eine offene Diskussionsrunde. Die Auswertung der Diskussion ergab, dass bei den Icons für „Personen zum Adressbuch hinzufügen“, „Markierung für einen bestimmten Ort auf einer Karte“ sowie „Bildschirm/Tablet drehen“ die vermuteten, gewünschten und von Google empfohlenen Icons übereinstimmen. Bei den Icons für „Navigation zu einem Ort“ und „Eingeschalteter Wecker/Alarm“ nannte hingegen keiner das empfohlene Icon. Dieses Ergebnis zeigt, dass für diese Bedeutungen andere Icons von der Zielgruppe gefordert werden. Abbildung 2 zeigt als erstes Icon das empfohlene und im Anschluss die vermuteten bzw. gewünschten Icons. Das Häkchen in dem empfohlenen Icon (Icon 4) sorgte in der Diskussion für Verwirrung und es kam die Frage auf, warum die Zeiger in diesem Icon verrutscht seien. Zudem wurden sich „Beine“ für den Wecker gewünscht, damit er eher wie ein realer, klassischer Wecker aussieht. Aufgrund der fehlenden „Beine“ wünschten die Teilnehmerinnen und Teilnehmer häufig die Glocke. Eine Teilnehmerin führte als Begründung für ihre Wahl (Icon 22) an: „das Plus bedeutet ‚an‘, ein Minus in der Glocke würde für ‚Wecker aus‘ stehen“. Während die Nennungen hier noch übereinstimmend waren, fielen die Nennungen für „Navigation zu einem Ort“ sehr divergent aus. Abbildung 3 zeigt empfohlene sowie vermutete bzw. gewünschte Icons.

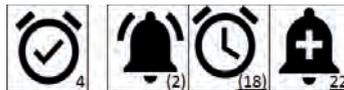


Abb. 2: Empfohlenes Icon, sowie (vermutete) und gewünschte Icons für „eingeschalteter Wecker/Alarm“

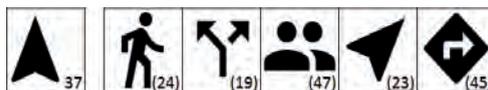


Abb. 3: Empfohlenes Icon, sowie (vermutete) und gewünschte Icons für „Navigation zu einem Ort“

Zur Beantwortung der Fragestellung „Welche Bedeutung hat das gegebene Icon?“ wurden den Teilnehmerinnen und Teilnehmern 9 Icons präsentiert, die jedem Smartphone-/ Tabletutzer nahezu täglich begegnen. Zunächst notierten die Teilnehmerinnen und Teilnehmer einzeln ihre vermuteten Bedeutungen für die gegebenen Icons. Im Anschluss erfolgte wieder eine offene Diskussion in der Gruppe. Gegeben waren die Icons, welche für „Senden“, „Aktualisieren“, „WLAN/Wifi“, „Anlage hinzufügen“, „Teilen“, „Bluetooth“, „Standort“, „verpasster Anruf“ und „Download“ empfohlen werden. Die erfahrenen Nutzer hatten lediglich bei den Icons für „Standort“ und „verpasster Anruf“ Probleme, die Gelegenheitsnutzer und Neulinge taten sich zusätzlich bei den Icons für „Senden“ und „Teilen“ schwer. Abbildung 4 zeigt die vier Icons, die zu Schwierigkeiten führten. Beim häufig verwendeten Icon für das „Absenden von Nachrichten“ (Icon 32) ist auffällig, dass Gelegenheitsnutzer und Neulinge bereits bei der ersten Aufgabe von der Nutzung dieses Icons abgeraten haben. Sie vermuteten Bedeutungen wie „Vergrößern“, weil es Ähnlichkeit mit dem mathematischen Größerzeichen hat, sowie nächste Seite, da es an einen Pfeil erinnert. Bei dem Icon für „Teilen“ (Icon 49) wurde die Bedeutung „Zwischenziele“ vorgeschlagen, da man mit einem Ziel anfängt und sich dieses dann in mehrere Teile aufteilt. Die erste Äußerung zum Icon für den Standort (Icon 50) war, „das ist doch das Zeichen von Tatort“, eine andere Vermutung war der Ausschaltknopf. Dass das Icon für verpasster Anruf (Icon 51) etwas mit Anrufen zu tun hat, haben alle Teilnehmerinnen und Teilnehmer an dem Telefonhörer erkannt, jedoch nicht, dass der Pfeil über dem Hörer für „verpasster Anruf“ steht. Die Vermutungen waren hier „telefonieren“, „Telefon an“, „Weiterleiten von Anrufen“, „Rückruf“. Die Verwendung dieser vier Icons sollte bei seniorentauglichen Apps überdacht werden.

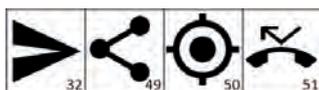


Abb. 4: Icons deren Bedeutung nicht eindeutig erkannt wurde

Die Vorstudie mit der Expertengruppe ergab einen weiteren Forschungsbedarf und den Bedarf an geeigneten Icons für „Eingeschalteter Wecker/Alarm“, „Navigation zu einem Ort“, „Versenden“, „Teilen“, „Standort“ sowie „verpasster Anruf“.

4 Ausblick

Für den nächsten Tablet-Workshop für Seniorinnen und Senioren wurde basierend auf den Antworten der Expertengruppe eine Auswahl missverständlicher Icons getroffen, sowie Variationen dieser erstellt. Diese Auswahl wird in einer Android-App verarbeitet, in der unter anderem Icons für die sechs oben genannten Bedeutungen ausgewählt werden können. Dadurch erfolgt eine Änderung des Studiendesigns. Es wird nicht mehr

mit Papier gearbeitet, sondern direkt in einer App. Dies wird unter anderem eine exaktere Auswertung der Häufigkeiten der gewählten Icons ermöglichen, da auf diese Weise z. B. auch die Meinungen von unsicheren Teilnehmerinnen und Teilnehmern erfasst werden können, die sich weniger an der Gruppendiskussion beteiligen. Zudem bietet dies eine weitere Möglichkeit die Interaktion mit den Tablets im Workshop zu trainieren. Dazu werden die Teilnehmerinnen und Teilnehmer einer Folgestudie im Mai 2017 jeweils zwischen sechs Icons für eine Bedeutung wählen können. Abbildung 5 zeigt die Icons die für „eingeschalteter Wecker/Alarm“ und „Versenden“ in der Folgestudie genutzt werden.



Abb. 5: Auswahl bestehender und eigener Icons (1. Reihe: „eingeschalteter Wecker/Alarm“; 2. Reihe: „versenden“). Schwarz: Google Material Icon; Lila: von anderen Designern erstelltes Icon im Stil des Material Design; Blau: eigene Icons

In mehreren Iterationen soll, mit einer größeren Stichprobe, ein für Seniorinnen und Senioren geeignetes Icon-Set entstehen, welches als Empfehlung für die Entwicklung von Senioren-Apps gelten soll. Zukünftig sollen auch weitere Informationen zum Hintergrund und Bildungsstand der Teilnehmerinnen und Teilnehmer erfasst werden. Für diejenigen Icons, für die kein eindeutiges und einheitliches Icon identifiziert werden kann, ist eine vorgeschaltete App bzw. Abfrage denkbar, welche es den Seniorinnen und Senioren ermöglicht, individuell ihr gewünschtes Icon für eine bestimmte Bedeutung zu wählen. Voraussetzung hierfür ist eine Unterstützung solcher Einstellungen in den einzelnen Apps.

Literaturverzeichnis

- [Ba16] Basenio, 2016, <https://www.basenio.de/senioren-ratgeber/technik/apps-fuer-senioren-die-besten-kostenlosen-apps-handynutzung-2016-471/>, Stand: 03.03.2017.
- [Er14] Erharter, D. et. al.: Smartphones, Tablets, App für Seniorinnen und Senioren. In (Kempter, G.; Ritter, W., Hrsg.): Assistenztechnik für betreutes Wohnen. Pabst Science Publishers, S. 221-235, 2014.
- [EX16] Erharter, D.; Xharo, E.: Developer-Guideline Usability von Apps für Seniorinnen und Senioren. Entstanden im Rahmen des Projektes „mobi.senior.A“, Online verfügbar unter <http://mobiseniora.at/app-entwicklung>, 2016.
- [Gf16] GfK Verein, 2016, <http://www.gfk-verein.org/compact/fokusthemen/nicht-ohne-mein-handy-auch-jenseits-der-50>, Stand: 03.03.2017.

- [LMG11] Leung, R.; McGrenere, J.; Graf, P.: Age-related differences in the initial usability of mobile device icons. *Behaviour & Information Technology*, S. 629-642, 2011.
- [Ma17] Material Design, <https://material.io/>, Stand: 03.03.2017.
- [ME16] MEEDIA, 2016, <http://meedia.de/2016/08/11/die-50-meistgenutzten-android-apps-deutschlands-pokemon-go-schon-in-der-top-ten/>, Stand: 03.03.2017.

Motiviert an die strukturierte Testfallermittlung

Doris Gabler¹, Martina Müller-Amthor², Georg Hagel³

Abstract: Im Rahmen einer Masterarbeit wird der Zusammenhang zwischen einer Online-Übung zur strukturierten Testfallermittlung, implementiert in einer Lehr-Lern-Plattform, und ihrer motivierenden Wirkung auf Informatik-Studierende untersucht. Dabei werden skalierte Lösungshinweise für das individuelle Einüben der strukturierten Testfallermittlung geliefert, die während der Präsenzlehre nicht im notwendigen Umfang geleistet werden können. Die Online-Übung führt dabei schrittweise an die strukturierte Bearbeitung der Testfallermittlung heran und bietet durch qualifiziertes Feedback Hilfestellungen beim Erlernen an. Erstmals im Sommersemester 2016 wurde die Online-Übung eng verzahnt mit der Präsenz-Übung eingesetzt und mit Lehrenden und Studierenden evaluiert. Letztere zeigten durch die Online-Übung eine höhere Motivation, sich mit dem Lerngegenstand auseinanderzusetzen. Für Lehrende vereinfacht das in der Online-Übung verwendete Schema die Einführung in das Thema.

Keywords: Lehre Software Engineering, Online-Übung, Testfallermittlung, Lehr-Lern-Plattform

1 Problemstellung

Die strukturierte Testfallermittlung wird den Informatik-Studierenden an der Hochschule Kempten im Modul des Basisstudiums „Grundlagen des Softwareengineerings“ zum Semesterende kurz vor den Prüfungen gelehrt. Das Modul bereitet durch die Lehre von abstrakten Methoden und Modellen [PSS09] auf komplexere Aufgaben im Studium vor [GI16]. Die strukturierte Testfallermittlung dient vor allem in umfangreichen Softwareentwicklungs-Projekten der systematischen Vorgehensweise in der Testphase. Da die Studierenden zu dem frühen Zeitpunkt im Studium meist erst mit kleineren Softwareentwicklungs-Projekten Erfahrung gesammelt haben, erkennen sie den Nutzen solcher abstrakten Themen noch nicht. Aufgrund dessen ist die Motivation der Studierenden gering, sich mit der strukturierten Testfallermittlung und ihrer umfangreichen Systematik auseinanderzusetzen, die wiederholt geübt werden sollte, um die Vorgehensweise zu verinnerlichen und Besonderheiten kennenzulernen [NDH08 bzw. Ort11]. Denn gerade im Software Engineering gilt, dass ein Auswendiglernen der Methoden nicht ausreicht, und eine Abarbeitung von Schritten ähnlich einem Kochrezept nicht möglich ist. Vielmehr muss situativ über jeden einzelnen Schritt entschieden werden [Ab12]. Diese Fähigkeit zur Analyse ist eine zentrale Anforderung an einen Software Engineer [Ja12], zugleich wurde diesbezüglich aber ein Defizit in der Informatik-Lehre identifiziert [RWK14].

¹ Hochschule Kempten, Fakultät Informatik, Bahnhofstraße 61, 87435 Kempten, doris.gabler@hs-kempten.de

² Hochschule Kempten, Fakultät Informatik, Bahnhofstraße 61, 87435 Kempten, martina.mueller-amthor@hs-kempten.de

³ Hochschule Kempten, Fakultät Informatik, Bahnhofstraße 61, 87435 Kempten, georg.hagel@hs-kempten.de

Die strukturierte Testfallermittlung wird an der Hochschule Kempten in einer je 90-minütigen Vorlesung und Übung in Präsenz behandelt, die zu wenig Zeit für ausreichende Übung lässt. Ein zusätzliches Angebot, selbstständig Übungen auf Papier zu bearbeiten und vom Übungsleiter korrigieren zu lassen, nutzten die Studierenden in der Vergangenheit kaum. Aus diesem Grund entstand die Idee, die Studierenden mit einer Online-Übung beim Erlernen anzuleiten und damit zu motivieren, die strukturierte Testfallermittlung zu üben und die Analyse-Fähigkeiten zu trainieren. Mediengestützte Lehr-Lern-Angebote im Bereich des Software Engineerings zielen bislang meist auf die automatische Bewertung von Programmierübungen ab [If14, Ki16, Pa16, We15]. Da es hier nicht um die Programmierung, sondern um die Analyse von Anforderungen geht, ist ein eigenes Tool notwendig - in Form eines anwendungsorientierten moodle-Plugins.

2 Strukturierte Testfallermittlung im lerntheoretischen Kontext

Zur Konzeption der Online-Übung wurde der Lerngegenstand der strukturierten Testfallermittlung im lerntheoretischen Kontext und im Zusammenhang mit der Zielgruppe untersucht. Das Modul „Grundlagen des Software Engineerings“ verfolgt insbesondere das Lernziel der Lösungskompetenz. Für die strukturierte Testfallermittlung als Prozesswissen setzt sich diese zusammen aus der Identifizierung und Übersetzung der notwendigen Informationen aus dem Anforderungstext, aus der Kenntnis und Einordnung von Fachbegriffen und aus der Wiedergabe und Anwendung der vorgegebenen Struktur. Letztere spiegelt sich in den Prozess-Schritten wieder, die die Studierenden nicht nur erklären, sondern auch durchführen können sollten, um damit die notwendigen Testfälle zu bestimmen:

Schritt 0: Bestimmung der Bereiche der gültigen und ungültigen Äquivalenzklassen

Schritt 1: Vergabe einer Äquivalenzklassen-ID

Schritt 2: Verfeinerung der Äquivalenzklassen

Schritt 3: Wahl eines Repräsentanten je Äquivalenzklasse

Schritt 4: Grenzwertanalyse

Schritt 5: Testfälle aus Grenzwerten

Schritt 6: Besondere Werte, die sich aus dem Anforderungstext ergeben

Neben den inhaltlichen Zielen ist es aufgrund der Motivations-Problematik wichtig, dass Anreize für die Studierenden geschaffen werden, die Online-Übung zu absolvieren und während der Übung nicht aufzugeben. Dies gestaltete sich mit den bisherigen Angeboten schwierig. Da bei der Bearbeitung auf Papier keine regelmäßige motivationsfördernde Interaktion stattfindet, fehlt den Studierenden die nötige Aufmerksamkeit und Konzentration [FHB13]. In der Konzeption der Online-Übung wurde also viel Wert auf motivationsfördernde Maßnahmen bei selbstgesteuertem Lernen gemäß dem ARCS-Prinzip gelegt [Or11 bzw. NDH08]:

- Aufmerksamkeit durch praxis- und prüfungsrelevante Aufgabenstellungen,
- Relevanz der studentischen Interaktion durch automatische Punktevergabe,

- Selbstvertrauen durch kleinschrittiges Feedback und regelmäßige Interaktion [KS08], um Überforderung zu vermeiden [Or11] und
- Zufriedenheit durch positives Feedback gemäß behavioristischer Ansätze [Ke13].

Die Umsetzung der strukturierten Testfallermittlung als Übung in einer Lehr-Lern-Plattform ermöglicht die Realisierung der hier genannten Anforderungen.

3 Umsetzung und Einsatz

An der Hochschule Kempten wird die Lehr-Lern-Plattform moodle zur Bereitstellung von Lern-Materialien genutzt, so dass die Studierenden damit vertraut sind. Allerdings ist kein Fragetyp verfügbar, mit dem sich die Prozess-Schritte wie gewünscht in einer Übungsaufgabe abbilden lassen. Darum wurde das für physikalische Aufgaben entwickelte Fragetyp-Plugin „formulas“⁴ angepasst, welches durch die Abfrage mehrerer Werte in Form eines Lückentextes, durch Variablen und durch formelbasierte Lösungsberechnung zumindest in Teilen dem konzipierten Fragetyp entspricht. Einige nicht benötigte Funktionalitäten wie die Abfrage von Maßeinheiten mit entsprechender Umrechnungsfunktion wurden deaktiviert, zusätzlich wurde das Plugin um die wahlweise Eingabe von Text oder Zahl erweitert. Das daraus entstandene Fragetyp-Plugin „findcases“ lässt eine schrittweise und kombinierte Auswertung der in die Lücken eingegebenen Werte pro Äquivalenzklasse zu, die vom Lernenden selbstständig nach Bedarf angestoßen werden kann.

Schritt: 1 2 3 4 5 6

ID	Definition	Repräsentant	Grenzwerte (GW)	Testfälle aus GW	besondere Werte	
sAq1	25000.01 ... MAX_FL	27000.00	25000.00 MAXm1	25.000.01 MAX_FL MAXp1	25.000.02 MAXm1 MAX_FL MAXp1	nv

Äquivalenzklasse

Feedback

Prüfen Sie Ihre Antwort:

1. Ist die Äquivalenzklassen-ID eindeutig?
2. Sind die genannten oberen / unteren Werte innerhalb (\leq bzw. \geq) oder außerhalb ($<$ bzw. $>$) der Äquivalenzklasse?
3. Entspricht der Repräsentant der Vorgabe "untere Grenze +2 \leq x \leq obere Grenze -2" und ist dabei kein besonderer Wert?
4. Sind die Grenzwerte wie folgt gewählt: untere Grenze / +1 / -1; obere Grenze / +1 / -1
5. Haben sie diejenigen Grenzwerte gestrichen, die in anderen Äquivalenzklassen geprüft werden sollten, sodass jeder Testfall aus den Grenzwertanalysen aller Äquivalenzklassen nur einmal vorkommt?
6. Gibt es besondere Werte oder haben Sie hier "nv" eingetragen?

Zensur für diese Einreichung: 0,00/5,00. Für diese Beantwortung erhielten Sie einen Punktabzug in Höhe von 0,50. Gesamtabzüge bisher: 1,50

Abb. 1: Vorab-Überprüfung einer Äquivalenzklasse (bei fehlerhafter Eingabe)

Möglichst einfache Syntax-Vorgaben gewährleisten, dass die Lösungen automatisch systemseitig ausgewertet werden können. Das Feedback pro Äquivalenzklasse gibt als Vorab-Überprüfung bei falscher Antwort die zu beachtenden Schritte und nicht einfach die richtige Lösung wieder. Somit werden die Studierenden angeregt, sich nochmals mit

⁴ https://moodle.org/plugins/qttype_formulas

ihrer eigenen Lösung auseinanderzusetzen und anhand der Hinweise ihre Fehler zu korrigieren.

Nach erstmaligem Einsatz im Sommersemester 2016 wurde das Lehr-Lern-Szenario im Wintersemester 2016/17 dahingehend verändert, dass schon vor der Präsenz-Übung statt der papierbasierten Übung die Online-Übung zum Selbststudium angeboten wurde. Die Übungsleiterin kann durch Analyse der bearbeiteten Aufgaben und ihrer Lösungswege in der Präsenz-Übung auf häufig auftretende Fehler und auf Probleme mit der strukturierten Testfallermittlung eingehen. Zusätzlich wird die Online-Übung ebenfalls in der Präsenz-Übung zur Erklärung und zur selbstständigen, individuellen Nutzung eingesetzt. Weitere Aufgaben stehen im Nachgang zur Prüfungsvorbereitung zur Verfügung.

4 Evaluation

Der Einsatz der Online-Übung im hier dargestellten Blended-Learning-Szenario erlaubt unterschiedliche Verfahren, um Motivation und Verstehen der Studierenden in Bezug auf die strukturierte Testfallermittlung zu evaluieren.

Durch Logfile-Analyse können die Schritte der strukturierten Testfallermittlung identifiziert werden, mit denen die Studierenden Schwierigkeiten haben. Wie im Vorfeld von den Dozierenden vermutet, liegen die Probleme hauptsächlich bei der Grenzwertanalyse. Außerdem lässt die Logfile-Analyse den Schluss auf große Unsicherheit in der strukturierten Testfallermittlung zu, weil die Studierenden die Vorab-Überprüfung häufig nutzen.

Beim Einsatz in der Präsenz-Übung kann die Entwicklung eines Wettstreits beobachtet werden, der unter den Studierenden durch den Vergleich der in der Online-Übung automatisch vergebenen Punkte entsteht. Die Online-Übung wirkt also als gamifizierendes Element für den Unterricht [Ke13]. Wenngleich sich die Studierenden durch die spielerische Konkurrenz begeistern lassen, kann man dennoch beobachten, dass sie sich aufgrund unterschiedlichen Vorwissens gemeinsam auf Fehlersuche begeben und dabei einander die strukturierte Testfallermittlung erklären.

Der im Sommersemester 2016 entwickelte Fragebogen dient dazu, die Haltung der Studierenden zur strukturierten Testfallermittlung zu eruieren. 26 der 29 Studierenden, die an der Evaluation im Sommersemester 2016 teilnahmen, geben an, sich durch die Online-Übung intensiver und motivierter damit beschäftigt zu haben. Die Vorgabe des Schemas durch die Online-Übung wird als Vorteil gesehen, weil dadurch das händische Anfertigen einer Tabelle für jede Übungsaufgabe entfallen und die Konzentration auf die eigentliche Aufgabe erfolgen kann. Studierende, die sich schon vorab intensiv mit dem Lerngegenstand auseinandergesetzt haben, werten dies allerdings eher als Nachteil, weil es das Lösungsschema in der Prüfung nicht gibt und die Online-Übung damit zu einfach ist. Besonders positiv wird die Möglichkeit der Vorab-Überprüfung gewertet (21 von 29), in der die Studierenden pro Äquivalenzklasse mit Hilfe von Hinweisen ihre Eingaben vor der endgültigen Abgabe korrigieren können, während die notwendigen

Syntax-Vorgaben und das Handling der Online-Übung allgemein wenige Probleme machen (9 von 29).

Gerade die systemseitig notwendigen Vorgaben und die damit einhergehende Komplexität der Online-Übung werden in einem Vorab-Interview von der Übungsleiterin kritisch gesehen. Diese Bedenken bestätigen sich nicht. Die Übungsleiterin kann die Online-Übung in der Präsenz sowohl für die Einführung in das Thema der strukturierten Testfallermittlung nutzen, als auch zur Motivation der Studierenden entsprechende Aufgaben selbstständig zu bearbeiten. Besonders positiv hervorzuheben ist, dass die Übungsleiterin in der Präsenz-Übung dadurch auf individuelle Fragen zum Thema eingehen und die Studierenden einzeln betreuen kann. Für die Übungsleiterin bedeutet dies, dass insgesamt in der Präsenz-Übung mehr Interaktion möglich ist, weil die Studierenden besser als durch papierbasierte Aufgaben aktiviert werden.

5 Fazit

Mit dem Einsatz der Online-Übung zur strukturierten Testfallermittlung können den Studierenden zwei Lernbedarfe erfüllt werden:

1. Die Studierenden bekommen mehr Übungsmöglichkeiten durch Bereitstellung unterschiedlicher Aufgaben zur strukturierten Testfallermittlung ohne Mehr-Aufwand bei der Übungsleitung, weil eine händische Bewertung der Lösungen entfällt.
2. Die Studierenden erhalten direktes und bedarfsorientiertes Feedback in kleinen Schritten, was einer möglichen Frustration beim selbstständigen Üben vorbeugt, und haben damit mehr Anreiz und Motivation konzentriert am Thema der strukturierten Testfallermittlung zu arbeiten.

Durch die Online-Übung werden sowohl für Studierende als auch für Dozierende positive Effekte erzielt. Diese werden auch in künftigen Semestern für die Lehre der strukturierten Testfallermittlung genutzt. Obwohl die Art der Aufgaben eher umfangreich sowie komplex ist und sich so von den in Online-Tests üblichen Lückentext- oder Multiple-Choice-Aufgaben unterscheidet, zeigt sich, dass sich diese Komplexität durch Hilfestellungen und Interaktionsmöglichkeiten in Online-Übungen beherrschen lässt.

Dieses Vorhaben wird aus Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 01PL12022C, 01PL17022C gefördert.

Literaturverzeichnis

- [Ab12] Abke, J.; Brune, P.; Haupt, W.; Hagel, G.; Landes, D.; Mottok, J.: EVELIN - ein Forschungsprojekt zur systematischen Verbesserung des Lernens von Software Engineering. In: Tagungsband Embedded Software Engineering Kongress 2012, Sindelfingen. Elektronikpraxis Würzburg; MicroConsult München, S. 653–658, 2012.

- [FHB13] Figas, P.; Hagel, G.; Bartel, A.: The Furtherance of Motivation in the Context of Teaching Software Engineering. In: 2013 IEEE Global Engineering Education Conference (EDUCON 2013). Berlin. IEEE, Piscataway, S. 1299–1304. Online verfügbar unter <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6530274>, Stand: 21.12.2016, 2013.
- [GI16] GI e. V. (Hrsg.): Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen. Online verfügbar unter https://www.gi.de/fileadmin/redaktion/empfehlungen/GI-Empfehlungen_Bachelor-Master-Informatik2016.pdf, Stand: 11.12.2016, 2016.
- [If14] Ifland, M.; Herrmann, F.; Ott, J.; Puppe, F.: WARP – ein Trainingssystem für UML-Aktivitätsdiagramme mit mehrschichtigem Feedback. In (Trahasch, S.; Plötzner, R.; Schneider, G.; Sassiati, D.; Gayer, C.; Wöhrle, N., Hrsg.): DeLFI 2014 – die 12. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. Freiburg. Köllen Bonn, S. 193-204, 2014.
- [Ja12] Janke, E.; Bartel, A.; Figas, P.; Brune, P.; Hagel, G.; Müller-Amthor, M.: Die Lehre von Software Engineering - eine Erhebung der Anforderungen aus der Praxis. In: Tagungsband Embedded Software Engineering Kongress 2012, Sindelfingen. Elektronikpraxis Würzburg; MicroConsult München, S. 653–658, 2012.
- [Ke13] Kerres, M.: Mediendidaktik. Konzeption und Entwicklung mediengestützter Lernangebote. 4., überarb. und aktualisierte Aufl. Oldenbourg, München, 2013.
- [Ki16] Kiesler, N.: Ein Bild sagt mehr als tausende Worte - interaktive Visualisierungen in web-basierten Programmieraufgaben. In (Lucke, U. et al., Hrsg.): DeLFI 2016 - die 14. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. Potsdam. Köllen Bonn, S. 339–341, 2016.
- [KS08] Kuhlmann, A. M.; Sauter, W.: Innovative Lernsysteme. Kompetenzentwicklung mit Blended Learning und Social Software. Springer-Verlag, Berlin, Heidelberg, 2008.
- [NDH08] Niegemann, H. M.; Domagk, S.; Hessel, S.: Kompendium multimediales Lernen. Springer, Dordrecht, 2008.
- [Or11] Ortmann-Welp, E.: Hybride Lernarrangements. Vernetzung von Präsenz- und Online-Lernen. Diplomica Verlag, Hamburg, 2011.
- [Pa16] Pape, S.; Flake, J.; Beckmann, A.; Jürjens, J.: STAGE. In (Dillon, L.; Visser, W.; Williams, L., Hrsg.): ICSE '16 38th International Conference on Software Engineering, Austin. The Association for Computing Machinery New York, S. 491–500, 2016.
- [PSS09] Pfannstiel, J.; Sängler, V.; Schmidt, C.: Game-based Learning im Bildungskontext einer Hochschule - ein Praxisbericht. In: MedienPädagogik - Zeitschrift für Theorie und Praxis in der Medienbildung (15/16), S. 1–21, 2009.
- [RWK14] Radermacher, A.; Walia, G.; Knudson, D.: Investigating the skill gap between graduating students and industry expectations. In (Jalote, P. et al., Hrsg.): ICSE '14 36th International Conference on Software Engineering, Hyderabad. The Association for Computing Machinery, New York, S. 291–300, 2014.
- [We15] Wegmann, C.: Erstellung eines Plug-ins für Unit-Tests für eine Lehr-/Lernplattform, unveröff. Bachelorarbeit im Stdg. Informatik an der Hochschule Kempten, 2015.

LearningView - ein digitales Werkzeug zur Unterstützung eines offenen Unterrichts

Michael Hielscher¹, Christof Tschudi², Doreen Prasse³, Beat Döbeli Honegger⁴

Abstract: Individualisierung und offene Unterrichtsformen erfordern umfangreiche Kompetenzen bei Lehrpersonen, aber auch bei Schülerinnen und Schülern, die ihren eigenen Lernprozess selbstständig steuern, planen und reflektieren sollen. Die zunehmende Verbreitung von mobilen Geräten (1:1-Ausstattung, BYOD) ermöglicht den Einbezug digitaler Werkzeuge zur Unterstützung von offenen Unterrichtsformen wie Wochen- und Arbeitsplänen. Gemeinsam mit Lehrpersonen wird an der Pädagogischen Hochschule Schwyz nach den Prinzipien des Design Based Research das Werkzeug *LearningView.org* entwickelt. In diesem Bericht stellen wir den aktuellen Stand und erste Erfahrungen an der Projektschule Goldau vor.

Keywords: offener Unterricht, Lernstandsdiagnose, digitale Werkzeuge

1 Einleitung

Die Fähigkeit selbstständig und eigenverantwortlich zu lernen gilt im Informationszeitalter als Schlüsselkompetenz und steht im Zentrum vieler Bildungsreformen der letzten Jahre (vgl. [FR14]). Um diese Kompetenzen bei Schülerinnen und Schülern zu fördern, werden offene, schülerzentrierte Unterrichtsformen postuliert, um auch den individuellen Bedürfnissen und Stärken der Kinder besser gerecht zu werden. Individualisierte Lehr- und Lernkonzepte gehen von der Grundannahme aus, dass jedes Kind unterschiedliche Begabungen, Interessen und Leistungsvoraussetzungen mitbringt, die es individuell zu fördern und fordern gilt. Bei all diesen Formen wird meist auch Wert auf die Selbstbestimmung der Schülerinnen und Schüler bei der Wahl ihrer Arbeitsaufträge und auf die Reflexion des eigenen Lernprozesses gelegt. Als Organisationsformen eines entsprechenden Unterrichts werden insbesondere Wochen- oder Arbeitspläne, Werkstatt- oder Stationen/Postenarbeit, Freiarbeit, sowie persönliche Kompetenzraster propagiert (vgl. [Pe03]). Für Lehrpersonen stellen solch offene Lehr- und Lernformen nicht nur eine inhaltliche, sondern auch eine organisatorische Herausforderung dar, gilt es doch sowohl mögliche Lernpfade von Schülerinnen und Schülern auszulegen als auch eine individualisierte Lerndiagnose und -beratung zu ermöglichen (vgl. [Ha03]).

¹ Pädagogische Hochschule Schwyz, IMS, Zaystrasse 42, CH-6410 Goldau, michael.hielscher@phsz.ch

² Gemeindeschulen Arth-Goldau, Schulhausplatz 1, CH-6410 Goldau, christof.tschudi@arth.educanet2.ch

³ Pädagogische Hochschule Schwyz, IMS, Zaystrasse 42, CH-6410 Goldau, doreen.prasse@phsz.ch

⁴ Pädagogische Hochschule Schwyz, IMS, Zaystrasse 42, CH-6410 Goldau, beat.doebeli@phsz.ch

2 Wochen- und Arbeitspläne als effektive Werkzeuge eines offenen Unterrichts

Bereits analoge Wochen- und Arbeitspläne bieten für Lehrpersonen rein organisatorisch eine handhabbare Möglichkeit, mehrere Lernende unterschiedlicher Lernstandniveaus zu betreuen, ohne den Überblick zu verlieren. In den selbstständigen Arbeitsphasen der Schülerinnen und Schüler ist die Lehrperson primär als Lernbegleiter oder Lerncoach tätig und kann gezielt auf einzelne Probleme und Fragen eingehen. Die kontinuierliche Überprüfung und Dokumentation erbrachter Leistungen erlaubt sowohl den Lernenden als auch der Lehrperson einen schnellen Überblick über den aktuellen Lern- und Arbeitsstand. Ob diese Möglichkeiten tatsächlich ausgeschöpft werden, hängt primär von den Kompetenzen der Lehrperson ab. Das betrifft beispielsweise die Auswahl und Gestaltung von Lernmaterialien und Lernumgebungen oder die angemessene individuelle Lernunterstützung für die Schülerinnen und Schüler (vgl. [Kr09]), die wiederum entsprechende diagnostische Fähigkeiten bei den Lehrpersonen voraussetzt (vgl. [Ha03] und [KW11]). Hascher sieht die Diagnose der individuellen Lernprozesse in einem individualisierten und offenen Unterricht aber nicht nur als Aufgabe der Lehrperson, sondern auch der Schülerinnen und Schüler, die ihren aktuellen Lernstand selbst einschätzen, diesen kommunizieren und ihren weiteren Lernprozess mitplanen. Die empirische Bildungsforschung attestiert der Einschätzung des eigenen Leistungsniveaus eine sehr hohe Wirksamkeit (vgl. [HBZ13]). Auch die Bereiche “formative Evaluation zur Diagnose” und “evidenzbasiertes Feedback durch die Lehrperson”, haben laut Hattie einen deutlich positiven Einfluss auf den Lernerfolg der Schülerinnen und Schülern.

3 Digitale Werkzeuge zur Unterstützung eines offenen Unterrichts

Offene Unterrichtsformen wie Wochenpläne und Lernjournale lassen sich mit Hilfe digitaler Werkzeuge abbilden und bieten - entsprechende Werkzeuge vorausgesetzt - Vorteile gegenüber klassischen papierbasierten Ansätzen. Insbesondere beim hohen zeitlichen und logistischen Aufwand der individuellen Lernstandsdiagnose können digitale Werkzeuge wertvolle Unterstützung bieten. So kann die Lehrperson, sowohl im Unterricht als auch während der Vorbereitung jederzeit den Leistungsstand, Lernprodukte oder Lernjournale der Lerngruppe einsehen und damit besser auf das unterschiedliche Lerntempo und Leistungsniveau einzelner Schülerinnen und Schüler reagieren. Umgekehrt können Schülerinnen und Schüler aufgrund der ständigen Verfügbarkeit ihrer gesamten Arbeitsstände und Selbsteinschätzungen, ihre eigenen Lernprozess gezielter steuern.

Zur digitalen Unterstützung offener Unterrichtsformen wurden in Schulen, insbesondere der Sekundarstufe II, anfänglich primär Lernplattformen wie Moodle, ILIAS, OLAT usw. eingesetzt. Diese Learning-Management-Systeme (LMS) wurden jedoch ursprünglich für den Hochschuleinsatz konzipiert, was sich auch in der Vielfalt der

Funktionen wie Einschreibungszeiträume, elektronische Prüfungen oder die umfangreichen Rollen- und Rechteverteilungen widerspiegelt. In der Regel werden diese für die gesamte Institution bereitgestellt und Lehrpersonen besitzen meist nur eingeschränkte Rechte. So werden häufig sowohl Kurse als auch Schülerkonti zentral erfasst und von einem Administrator verwaltet. Die Systeme bieten den Lehrpersonen meist vielfältige Möglichkeiten, Unterrichtsmaterialien und Aufträge für die Lernenden bereitzustellen. Diverse Lernaktivitäten wie Quizzes, Umfragen oder Dateiabgaben ermöglichen in der Regel auch tabellarische Übersichten für die Lehrperson darüber, welche Aufgaben bereits von welchem Lernenden erledigt und welche Punktzahl erreicht wurde. Diese Werkzeuge und Übersichten können damit potentiell zur Lernstandsdiagnostik herangezogen werden. Sie stellen zudem die auf diesem Weg erfassten, digitalen Schülerprodukte jederzeit zur Einsicht bereit. Eine größere Verbreitung haben LMS bisher jedoch nicht gefunden. Das gilt insbesondere auf den unteren Schulstufen. Dies ist vermutlich auf zwei Gründe zurückzuführen. Einerseits standen bisher auf der Primarstufe Computer meist nur in sehr begrenzter Anzahl im Klassenzimmer zur Verfügung und digitale Werkzeuge wurden primär von den Lehrpersonen genutzt. Andererseits ist dies vermutlich der Komplexität der Werkzeuge sowohl in der Bedienung für Lehrpersonen und Schülerinnen und Schüler als auch in der Administration geschuldet.

Neben den institutionell zur Verfügung gestellten Lernplattformen (LMS) existieren inzwischen viele kleinere und spezifischere Apps und Dienste, die einen offenen und individualisierten Unterricht unterstützen können. Das MFID-Modell (vgl. [RH17]) bietet eine umfangreiche Zusammenstellung von digitalen Werkzeugen zur Diagnose und Klassenführung. So ermöglichen z.B. Dienste wie Google Classroom, showbie.com oder handouts.in die individualisierte Verteilung digitaler Arbeitsblätter und Aufträge und die anschließende Korrektur und Bewertung durch die Lehrperson. Werkzeuge wie seesaw.me ermöglichen hingegen den Schülerinnen und Schülern digitale Lernprodukte in einem persönlichen Portfolio zu sammeln und der Lehrperson, der Klasse oder den Eltern zur Einsicht zu stellen. Diese Werkzeuge sind im Vergleich zu LMS in der Handhabung meist deutlich einfacher, bilden jedoch immer nur einen Teil des Unterrichts ab. Eine Gesamtübersicht über den Arbeitsstand der Kinder (inklusive Aufgaben aus Lehrmitteln, Projektarbeiten usw.) ist mit einer solchen Lösung, oder auch mit Kombinationen mehrerer Lösungen, kaum möglich.

In den letzten Jahren haben 1:1-Ausstattungen insbesondere mit Tablets sowie Bring-Your-Own-Device-Projekte (BYOD) in der Primarschule deutlich zugenommen (vgl. [BA17]). Diese Verfügbarkeit persönlicher und rasch einsatzfähiger Geräte erhöht das Potenzial der Unterstützung von individuellen Lernprozessen durch digitale Planungs- und Dokumentationsinstrumente. Um sowohl den besonderen Potenzialen mobiler Geräte gerecht zu werden, als auch um die stufenspezifischen Bedürfnisse von Primarlehrpersonen bereits in der Werkzeugentwicklung einfließen zu lassen, entwickeln wir in einem explorativ orientierten Forschungsprojekt (*Design Based Research*) eine Lernumgebung für persönliche Endgeräte (Smartphones, Tablets). Ziel ist es, eine primarschultaugliche Lösung zu finden, um Lern- und Arbeitsstände durch die

Schülerinnen und Schüler selbst zu erfassen, Selbsteinschätzungen und Reflexionen in einem Lernjournal zu dokumentieren und digitale Arbeitsprodukte abzulegen.

In einem Forschungs- und Entwicklungsprojekt wurde das Werkzeug *LearningView* nach dem Ansatz des Design Science Research zusammen mit Lehrpersonen entwickelt. *LearningView* ist für Schülerinnen und Schüler ein Werkzeug zur Steuerung, Dokumentation und Reflexion ihres eigenen Lernprozesses und zur Förderung ihrer Selbstlernkompetenzen (siehe Abb. 1). Lernende wählen dabei selbstständig Aufgaben aus einem Aufgabenpool aus und markieren eigenverantwortlich deren Lösung. Die Lehrperson gibt Hinweise zu Anforderungsniveau, Bearbeitungszeit, Sozialform und weiteren Rahmenbedingungen, um die Lernenden bei der Auswahl von Aufgaben zu unterstützen. Lernaktivitäten können per Kamera und Mikrofon des mobilen Schülergeräts festgehalten und in einem Lernjournal dokumentiert werden. Zu jedem Thema können sowohl die Schülerinnen und Schüler eine Selbst- als auch die Lehrperson eine Fremdbeurteilung vornehmen.

Für Lehrerinnen und Lehrer ist *LearningView* primär ein organisatorisches und diagnostisches Werkzeug für einen individualisierten Unterricht. Die Lehrperson stellt sich ihren eigenen Materialienpool zusammen, den sie auch mit Kollegen teilen kann. Sie wählt anschliessend für jede Klasse oder individuell für jede Schülerin passende Materialien aus ihrem Pool aus (siehe Abb. 2). Eine tabellarische Übersicht über den Arbeitsstand der Klasse und eine chronologische Aktivitätsanzeige unterstützen die Lehrperson bei der Lernstandsdiagnose und geben Hinweise, wo Unterstützung nötig ist.

Weitere Informationen finden sich auf der Projektwebseite: <https://learningview.org>



Abb. 1: LearningView-Ansicht für Schülerinnen und Schüler

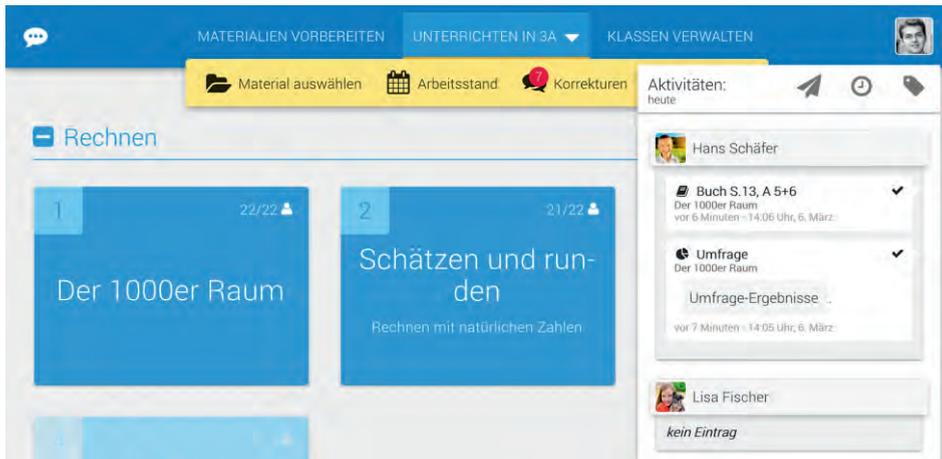


Abb. 2: LearningView-Ansicht für die Lehrperson

4 Erfahrungen in der Projektschule Goldau

Das Projekt wird in enger Zusammenarbeit mit der Projektschule Goldau entwickelt. An dieser Primarschule arbeiten die Schülerinnen und Schüler seit 2009 mit persönlichen mobilen Geräten (vgl. [DN10]), seit 2013 in einem BYOD-Setting. Seit Mitte 2015 wurde *LearningView* mit ausgewählten Schulklassen (5./6. Klasse) erprobt. In den Versuchsklassen wurde das Werkzeug primär im Mathematikunterricht und teilweise im Fach Natur-Mensch-Gesellschaft eingesetzt. Die bisherigen Arbeitspläne auf Papier wurden inzwischen durch das digitale Werkzeug abgelöst. Für die Lehrperson besteht eine der größten Veränderungen des Unterrichts darin, dass Schülerinnen und Schüler mit *LearningView* jetzt immer wissen, was sie als nächstes machen könnten. Zudem schätzen die Lernenden die klaren Zielvorgaben bezüglich des Anforderungsniveaus der einzelnen Aufgaben. *LearningView* konnte sich aus der Sicht der Lehrpersonen sowie der Schülerinnen und Schüler innerhalb kurzer Zeit als ganz alltägliches Werkzeug im Unterricht etablieren. Dabei zeigen die Zugriffsstatistiken, dass *LearningView* auch außerhalb der Unterrichtszeiten regelmäßig für die Erledigung von Hausaufgaben oder für Selbsteinschätzungen genutzt wird. *LearningView* wird von den Lehrpersonen als effizienzsteigernd erlebt, da der Aufwand zur Erteilung von Arbeitsaufträgen relativ gering ausfällt. Die Möglichkeiten zur Aufzeichnung und Einreichung von Audios und Fotos, sowie der asynchrone und synchrone Einsatz von Umfragen eröffnen zudem attraktive Unterrichtsszenarien. Ob und inwiefern sich die Selbstlernkompetenzen von Schülerinnen und Schüler durch den Einsatz des Werkzeugs verbessern, bleibt Gegenstand zukünftiger Untersuchungen.

5 Fazit und Ausblick

Die zunehmende Verbreitung von 1:1 ausgestatteten Klassenzimmern wirft Fragen zum didaktisch sinnvollen Einsatz digitaler Geräte in einem individualisierten und schülerzentrierten Unterricht auf. Das vorgestellte Werkzeug *LearningView* bietet sowohl vielfältige Möglichkeiten für Schülerinnen und Schülern ihren Lernprozess eigenverantwortlich zu steuern und zu reflektieren als auch diagnostische und organisatorische Hilfsmittel für Lehrpersonen. *LearningView* verzichtet bewusst auf eine umfangreiche Vorstrukturierung und lässt damit der Lehrperson viel Freiraum. So wäre auch der vollständige Verzicht auf Aufgaben und die reine Nutzung für Kompetenzraster denkbar. Um die Potentiale und Auswirkungen des Werkzeugs für bzw. auf den Unterricht genauer zu analysieren, werden ab Sommer 2017 weitere Schulklassen einbezogen und mit Interviews und Fragebögen begleitet. Der Einsatz auf höheren Schulstufen soll zukünftig ebenfalls evaluiert werden. Für den Einsatz im Hochschulumfeld und der Erwachsenenbildung ist *LearningView* hingegen nicht konzipiert. Auch die Frage, ob und inwiefern durch die Verwendung von *LearningView* die Selbstlernkompetenzen von Schülerinnen und Schülern verbessern, ist bisher ungeklärt und soll zu einem späteren Zeitpunkt untersucht werden. *LearningView* steht interessierten Lehrpersonen über die Webseite <http://learningview.org> und die App-Stores kostenlos zu Testzwecken zur Verfügung.

Literaturverzeichnis

- [BA17] Bastian, J.; Aufenanger, S.: Tablets in Schule und Unterricht: Forschungsmethoden undperspektiven zum Einsatz digitaler Medien. Springer-Verlag, 2017.
- [DN10] Döbeli Honegger, B.; Neff, C.: Personal Smartphones in Primary School - Devices for a PLE? International Journal of Virtual and Personal Learning Environments Vol. 2/04 S. 40-48, 2010.
- [FR14] Fischer, C.; Rott, D.: Individuelle Förderung als schulische Herausforderung. Friedrich-Ebert-Stiftung, 2014.
- [Ha03] Hascher, T.: Diagnose als Voraussetzung für gelingende Lernprozesse. Journal für Lehrerinnen- und Lehrerbildung 2, S. 25-30, 2003.
- [HBZ13] Hattie, J.; Beywl, W.; Zierer, K.: Lernen sichtbar machen. Schneider-Verlag Hohengehren, 2013.
- [KW11] Klieme, E.; Warwas, J.: Konzepte der individuellen Förderung. Zeitschrift für Pädagogik, 57/06, S. 805-818, 2011.
- [Kr09] Krammer, K.: Individuelle Lernunterstützung in Schülerarbeitsphasen. Münster: Waxmann, 2009.
- [Pe03] Peschel, F.: Offener Unterricht. Schneider Verlag Hohengehren, 2003.
- [RH17] Rodemer, T.; Hamsch, J.: Digitalisierung an Schulen: MIFD in der Praxis: Diagnose und Selbsteinschätzung, L.A. Multimedia, Ausgabe Februar Heft 1, 2017.

Vorgehensmodell zur Identifikation, Aufnahme und Aufbereitung von Prozesswissen in der Industrie 4.0

Niklas Kreggenfeld¹, Christopher Prinz², Carsten Ullrich³ und Bernd Kuhlenkötter⁴

Abstract: Der stetig wachsende Wettbewerbsdruck auf das produzierende Gewerbe in Deutschland fordert einen hohen Einsatz der Industrie, um den Produktionsstandort Deutschland zu sichern. Diesem Wettbewerbsdruck begegnen deutsche Unternehmen unter anderem mit hohen Qualitätsstandards sowie einem hohen Automatisierungsgrad, um gegen geringere Lohnkosten der Wettbewerbsländer bestehen zu können. Gleichzeitig wird die deutsche Produktionswirtschaft vom Demografischen Wandel unter Druck gesetzt. Um zum einen dem immer steigenden komplexen Produktionsumfeld und dem demografischen Wandel entgegenzuwirken, sind Assistenzsysteme wertvolle Hilfsmittel. Diese ermöglichen Unternehmen ihre Mitarbeiter viel flexibler einzusetzen und zum anderen stellen sie für die sich wandelnde Belegschaft auf dem Shopfloor eine Möglichkeit des lebenslangen Lernens dar. Im folgenden Beitrag wird auf die Erstellung von Inhalten als Form von strukturiert aufbereitetem Prozesswissen für derartige Assistenzsysteme eingegangen.

Keywords: Assistenzsysteme, Industrie 4.0, Prozessvisualisierung

1 Einleitung

Der Begriff „Industrie 4.0“ beschreibt im Wesentlichen die technische Integration von Cyber-Physischen-Systemen (CPS) in Produktion und Logistik sowie die damit verbundene Anwendung des Internets der Dinge auf industrielle Prozesse, sodass Cyber-Physische Produktionssysteme (CPPS) entstehen [KWH13]. So soll es den Unternehmen ermöglicht werden, weltweit schneller und flexibler auf Kundenanforderungen sowie externe und interne Einflüsse zu reagieren [Dr14]. Anders als noch beim CIM-Ansatz (Computer-Integrated Manufacturing) aus den 1980er Jahren wird im Kontext der Industrie 4.0 nicht die menschenleere Fabrik angestrebt. Ganz im Gegenteil arbeiten trotz der fortschreitenden Automatisierung Mensch und Maschine in CPPS eng verzahnt zusammen und bilden somit ein sozio-technisches System. Neben den angestrebten positiven Entwicklungen durch CPPS stellen diese Unternehmen jedoch auch zunehmend vor Probleme, da die Komplexität technischer Systeme im Produktionsumfeld enorm ansteigt und somit ein rasanter Anstieg des Informationsbedarfs sowie des notwendigen beruflichen Kompetenz- sowie Expertiseniveaus des Produktionspersonals entsteht

¹ Lehrstuhl für Produktionssysteme, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, kreggenfeld@lps.rub.de

² Lehrstuhl für Produktionssysteme, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, prinz@lps.rub.de

³ Educational Technology Lab (EdTec), DFKI GmbH, Alt-Moabit 91c, 10559 Berlin, carsten.ullrich@dfki.de

⁴ Lehrstuhl für Produktionssysteme, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, kuhlenkoetter@lps.rub.de

[Kr14]. Darüber hinaus führt dieser demografische Wandel zu einer Verknappung des Arbeitskräfteangebots, woraus vermehrt die Einstellung geringer qualifizierten Personals resultiert [Kr14], [Ka12]. Am Ende entsteht in sämtlichen Lebenszyklusphasen von Maschinen und Anlagen eine deutliche Schere zwischen der Anlagen- und Aufgabenkomplexität einerseits sowie der Befähigung des Personals andererseits (Abbildung 1).

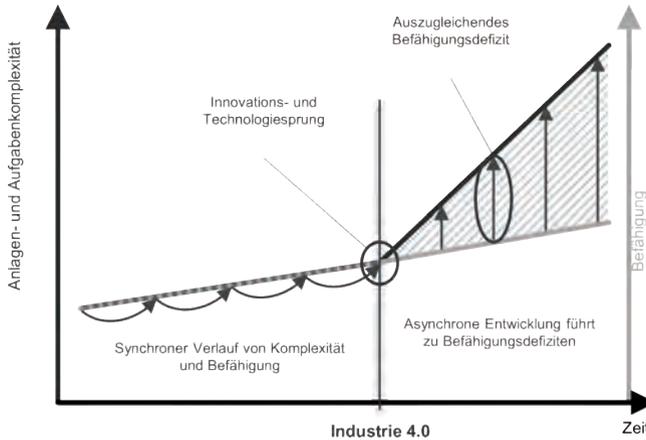


Abb. 1: Die durch Industrie 4.0 steigende Komplexität am Arbeitsplatz führt zu einem Befähigungsdefizit [Kr14]

Insbesondere stellen hierbei Nicht-Routine-Tätigkeiten eine enorme Herausforderung dar. Unter Nicht-Routine-Tätigkeiten werden in diesem Zusammenhang Tätigkeiten verstanden, bei denen eine geringe Routinebildung angenommen werden kann. Dieser Sachverhalt impliziert die Notwendigkeit innovativer Assistenzsysteme, die Produktionsmitarbeiter jeglichen Alters, sozialen Hintergrunds und Kompetenzniveaus bei Tätigkeiten (z. B. Instandhaltung, Bedienung, Inbetriebnahme) an hochkomplexen Anlagen durch Bereitstellung entsprechender Wissens- und Assistenzdienste unterstützen. Im Forschungsprojekt APPSist wurde ein solches Assistenzsystem systematisch entwickelt.

2 Das APPSist-Assistenzsystem

Das im Forschungsprojekt APPSist entwickelte Assistenz- und Lernsystem stellt eine neue Generation mobiler, kontextsensitiver und intelligent-adaptiver Assistenzsysteme für die Bereitstellung von Wissen und die Handlungsunterstützung in der Industrie 4.0 dar. Das System soll dabei Befähigungsdefizite bei der Durchführung von Nicht-Routine-Tätigkeiten auf dem Shopfloor im Sinne eines Performance Supports ausgleichen. Dieser Support kann bei Tätigkeiten der Inbetriebnahme, Bedienung, Wartung, Instandhaltung,

aber auch der Montage erfolgen. Ergänzend dazu können Mitarbeiter nachhaltig Wissen durch die Bereitstellung von kleineren didaktisch aufbereiteten Lerneinheiten („Learning Nuggets“) erlangen. Ziel ist es, die berufliche Entwicklung der Mitarbeiter zu fördern, damit diese zukünftig anspruchsvollere Tätigkeiten durchführen können. Aus rein technischer Sicht betrachtet ist das APPSist-System als diensteorientierte Architektur aufgebaut. Um Wissen einzelnen Komponenten des Shopfloors (z. B. Bereiche, Stationen, Maschinen, Personen) zuordnen zu können, verfügt das System über eine Ontologie, welche die Komponenten und Wissens-/Assistenzinhalte miteinander Verknüpft [U115], [U116].

3 Vorgehensmodell zur Assistenzprozesserstellung

Um Prozesse für die Unterstützung von Mitarbeitern bei der Durchführung von Nicht-Routine-Tätigkeiten auf dem Shopfloor zu assistieren, bedarf es einer strukturierten Vorgehensweise, um das entsprechende im Unternehmen vorhandene Prozesswissen zu identifizieren, aufzunehmen, zu optimieren und entsprechend aufzubereiten. Hierzu wurde im Forschungsprojekt APPSist ein Vorgehensmodell entwickelt, welches im Folgenden vorgestellt werden soll.

3.1 Identifikation von notwendigen Prozessen

Die Auswahl von zu assistierende Prozessen stellt die erste Herausforderung dar und wird mithilfe verschiedener Kriterien durchgeführt. Zu diesen Kriterien zählen u.a.:

- unternehmens- oder bereichsspezifische Kennzahlen, wie z. B. die Meantime-To-Repair (MTTR), das Liefertreuerisiko oder die Anlagenverfügbarkeit, die ein Unternehmen verbessern möchte;
- Einsatzmöglichkeit von Mitarbeitern;
- Wissensverlustrisiko (Wegfall von Wissensträgern, geringe Wissensdurchdringung).

Ebenso können weitere Kriterien durch ein Unternehmen selbst aufgestellt werden, die sie bei der Auswahl der relevanten Prozesse unterstützen.

3.2 Ausführungsanalyse

Die so ermittelten Prozesse müssen anschließend systematisch aufgenommen und für die Verwendung im APPSist-System aufbereitet werden. Die Modellierung erfolgt in BPMN (Business Process Model and Notation). Hierzu wird zunächst eine Ausführungsanalyse durchgeführt, das heißt, der Prozess wird von Experten im Unternehmen so definiert, wie er bislang Anwendung findet (subjektive Arbeitsweise). Hierbei können einzelne Prozessschritte weiter ausdetailliert werden („Aktivitäten“), sodass eine mehrstufige

Prozessstruktur entsteht (Abbildung 2). Schritte, die nicht detailliert werden müssen, werden als „Elementarschritte“ bezeichnet.

Um eine hohe Qualität der Dokumentation des Prozesses sicherzustellen, ist eine Reihe von Kriterien zu beachten, wie der Grundsatz der Richtigkeit, der Relevanz, der Wirtschaftlichkeit, der Klarheit, der Vergleichbarkeit sowie der Objektivität bzw. Reliabilität [BPV12].

3.3 Festlegung der Arbeitsmethode

In einem zweiten Schritt wird der Prozess überarbeitet und optimiert, um einen „Best-Practice-Ansatz“ zu generieren (*Festlegung einer definierten Arbeitsmethode*). Die Optimierung kann durch Eliminieren, Standardisieren, Substituieren, Integrieren, Vermeidung von Iterationen und Beschleunigen erfolgen [Sc11]. Da die modellierten Prozesse zur systematischen Unterstützung von Mitarbeitern dienen sollen, müssen sie neben den oben genannten Kriterien weiterhin eindeutig und nachvollziehbar, reproduzierbar sowie vollständig (d.h. sie müssen alle möglichen auftretenden Schritte enthalten) sein.

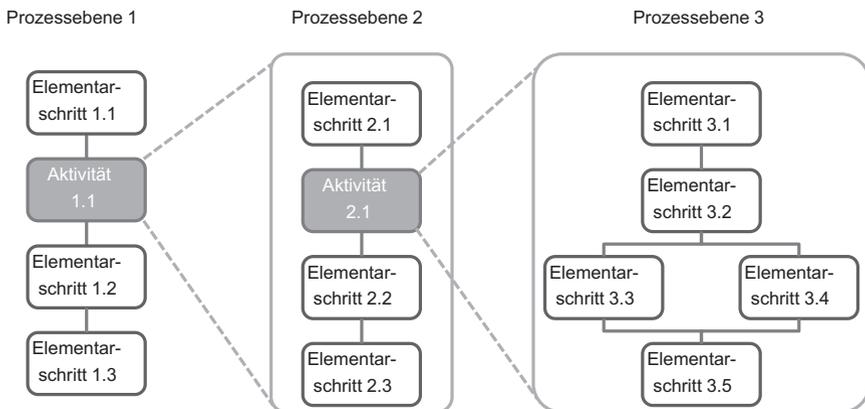


Abb. 2: Prozessstruktur mit mehreren Prozessebenen

Entscheidend für die Adaptivität bei der Nutzung eines Assistenzprozesses sind die Aktivitäten. Diese stellen Entscheidungspunkte dar, an denen das Assistenzsystem, je nach Expertise des Mitarbeiters, entweder den Prozess im Detail über verschiedene Ebenen hinweg anzeigt (in Abbildung 2 beispielsweise zusätzlich Ebene 2 statt nur Ebene 1) oder den Prozess weniger detailliert (z. B. auf Prozessebene 1) durchläuft. Hierbei ist zu beachten, dass ein Prozess so wenig detailliert wie möglich, aber so feingranular wie nötig bestimmt werden muss. Zur Festlegung der „richtigen“ Granularität ist es hilfreich, die angestrebte Zielgruppe des Prozesses zu betrachten.

3.4 Prozessaufbereitung

Nach Festlegung der Arbeitsmethode müssen in jedem Schritt die im Assistenzsystem anzuzeigenden Inhalte sowie Systemschnittpunkte und -rückkopplungen festgelegt werden. Die vollständige Beschreibung eines Prozessschrittes besteht letztendlich aus der Überschrift des Schrittes, einem verständlichen Anweisungstext, bei Bedarf einem besonderen Warn- oder Hinweistext, erklärenden Medien (Fotos, Videos, AR etc.) sowie einer Zuordnung von Schnittpunkten und Daten zum Triggern des Prozesses sowie zur Ausführungskontrolle während der Prozessdurchführung.

3.5 Prozessvalidierung

Bevor der Prozess in das APPSist-System integriert wird, ist es unabdingbar, diesen durch eine Expertengruppe zu validieren. Hierzu ist es ratsam, den Prozess systemnah abzubilden, wozu sich beispielsweise einfache PowerPoint-Folien im APPSist-System-Design eignen. Hier werden alle zuvor bestimmten Inhalte eingepflegt, sodass ein anschauliches Prüfdokument entsteht. Dieses dient zur Überprüfung der Prozessstruktur sowie von Sicherheitsaspekten (Sicherheitshinweise und Symbole) und Assistenzinhalten (Art der Darstellung, Korrektheit, Verständlichkeit).

Besteht bei einem dieser Punkte noch Optimierungsbedarf, so müssen der Prozess und/oder die Inhalte nochmals überarbeitet werden. Erst nach Freigabe aller Aspekte durch die Experten, kann der Assistenzprozess im Assistenzsystem verwendet werden.

4 Zusammenfassung

Das vorgestellte Prozesserkonzept ermöglicht eine strukturierte und systematische Vorgehensweise zur Identifikation, Strukturierung, Erfassung, Digitalisierung, Visualisierung und Nutzbarmachung von Prozesswissen. Die digitale Speicherung und Darstellung der Prozesse in Form der Business Process Model and Notation Sprache bietet dazu eine gute Möglichkeit, komplexe Prozesse nach festen Gestaltungsregeln zu hinterlegen. Es konnte gezeigt werden, dass die Erstellung nach wie vor sehr aufwendig ist, aber mithilfe der im Projekt APPSist erarbeiteten Vorgehensweise Rahmenbedingungen entwickelt wurden, um die Erstellung von Assistenzprozessen strukturiert durchzuführen. Durch die systematische Aufnahme der Prozesse erfolgt eine nachhaltige Speicherung von Prozesswissen im Unternehmen. Gleichzeitig erfolgt durch die Assistenz auf Basis dieser Prozesse auch eine Verteilung des Wissens, welches die Mitarbeiter somit on-the-job erlangen können.

5 Förderhinweis

Dieser Beitrag entstand im Rahmen des Projekts „APPSist – Intelligente Wissensdienste

für die Smart Production“, das vom Bundesministerium für Wirtschaft und Energie unter dem Kennzeichen 01MA13004C gefördert und vom DLR-Projektträger betreut wird.

Literaturverzeichnis

- [BPV12] Becker, J.; Probandt, W.; Vering, O.: Grundsätze ordnungsmäßiger Modellierung. Konzeption und Praxisbeispiel für ein effizientes Prozessmanagement. Springer Gabler, Berlin Heidelberg, 2012.
- [Dr14] Drapp, B.: Industrie 4.0 – die Chance für den Mittelstand. In: (Garn, M.; Schleidt, D., Hrsg.): Jahrbuch Innovation 2014. Innovationstreiber für Wirtschaft, Wissenschaft, Politik und Gesellschaft. FAZ-Institut, Frankfurt am Main, S. 80–81, 2014.
- [KWH13] Kagermann, H.; Wahlster, W.; Helbig, J.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Abschlussbericht des Arbeitskreises Industrie 4.0. Deutschlands Zukunft als Produktionsstandort sichern. 2013.
- [Ka12] Kay, R.: Demographischer Wandel: personalpolitische Herausforderungen, Problembewusstsein und Anpassungsstrategien von KMU. In: (Charta der Vielfalt e.V., Hrsg.): Jung - Alt - Bunt. Diversity und der demographische Wandel. Berlin, S. 22–24, 2012.
- [Kr14] Kreimeier, D. et al.: Intelligente Wissensdienste in Cyber-Physischen Systemen. Industrie Management 6, S. 25–29, 2014.
- [Sc11] Schuh, G. et al.: Prozessmanagement. In: (Schuh, G.; Kampker, A., Hrsg.): Strategie und Management produzierender Unternehmen. Handbuch Produktion und Management. Springer, Berlin Heidelberg, S. 327–382, 2011.
- [U115] Ullrich, C. et al.: Assistenz- und Wissensdienste für den Shopfloor. In: (Rathmayer, S.; Pongratz, H., Hrsg.): Proceedings der Pre - Conference Workshops der 13. E - Learning Fachtagung Informatik - DeLFI 2015. München, S. 47–55, 2015.
- [U116] Ullrich, C.: An Ontology for Learning Services on the Shop Floor. In: (Sampson, D.G.; Spector, S.M.; Ifenthaler, S.; Isa, P., Hrsg.): Proceedings of 13th International Conference on Cognition and Exploratory Learning in Digital Age. IADIS Press, S. 17–24, 2016.

Entwicklung einer Evaluation für Blended Learning Konzepte

Kristin Vogelsang¹, Ilse Hagerer², Uwe Hoppe³ und Kirsten Liere-Netheler⁴

Abstract: Lehrevaluationen haben sich als adäquates Mittel, um Lehrqualität messbar zu machen und kontinuierlich verbessern zu können, etabliert. Sie sind meist auf das Dozentenverhalten und die Interaktion zwischen Studierenden und Dozenten ausgerichtet. In digital unterstützten Blended Learning Situationen jedoch, in denen der Lehrende nicht allein im Fokus der Studierenden steht, gilt es, andere Maßstäbe anzuwenden, um formativ mittels Lehrevaluation die Wissensvermittlung erfolgreich zu gestalten. Evaluationswerkzeuge müssen stärker auf die Vielschichtigkeit und die besonderen Aspekte des Blended Learning eingehen. In diesem Beitrag wird ein Instrument entwickelt, welches ermöglicht, Blended Learning Veranstaltungen zu evaluieren und dabei vor allem auf die bisher wenig betrachteten Bereiche des Lernerfolges und der Technologie-Akzeptanz seitens der Studierenden einzugehen. Anhand einer beispielhaften Masterveranstaltung wird die Evaluationsmethodik entwickelt und analysiert. Zudem wird die Vorteilhaftigkeit theoriegeleiteter Verfahren aufgezeigt.

Keywords: Blended Learning, Evaluation, Technologie-Akzeptanz, Fallstudie, Digitalisierte Wissensvermittlung

1 Einleitung

Die zunehmende Digitalisierung der Gesellschaft wird verstärkt auch Bestandteil des Lehr- und Lernalltags an Universitäten. Die Möglichkeiten der Wissensvermittlung bewegen sich zwischen klassischer Präsenzlehre ohne Medieneinsatz und vollständig internetbasierten Vorlesungen [Ru14]. Dabei ist es mit Blick auf den Lernerfolg wichtig, die Lehre an den individuellen Wissenserwerb der Studierenden anzupassen bzw. anpassbar zu gestalten [B100]. Dies gelingt mit Hilfe eines wohlbedachten Technologieeinsatzes. Der Medieneinsatz ist dann vorteilhaft, wenn er dazu genutzt wird, Wissen zu vermitteln und den Lernalltag der Studierenden zu verbessern [Gr06]. Damit öffnet sich jedoch ein weiterer Bereich, der Einfluss auf die Qualitätswahrnehmung der Studierenden sowie auf die Lehre nimmt. Aus Gründen der Qualitätssicherung ist es wichtig, diese Veränderung des Lernalltags im Rahmen von Lehrevaluationen zu analysieren. Obwohl zahlreiche Varianten der Evaluation von Lehrveranstaltungen existieren, besteht das Problem der Datenerhebung, wenn Evaluation nicht primär auf die Qualität des Lehrenden fokussiert und die Ebene

¹ Universität Osnabrück Katharinenstraße 1 49069 Osnabrück, kristin.vogelsang@uos.de

² Universität Osnabrück Katharinenstraße 1 49069 Osnabrück, ilse.hagerer@uos.de

³ Universität Osnabrück Katharinenstraße 1 49069 Osnabrück, uwe.hoppe@uos.de

⁴ Universität Osnabrück Katharinenstraße 1 49069 Osnabrück, kirsten.liere-netheler@uos.de

der medialen Wissensvermittlung hinzukommt. Ein einheitliches Verfahren der Lehrevaluation im Blended Learning existiert nicht [GKV06]. Die systematische Analyse des Umgangs mit der Technik, der Integration in den Lernalltag und der auf diese Weise erlernten Kompetenzen sind jedoch notwendig [Kr01, SSR16, TDP11].

Ziel des Artikels ist es, ein Modell vorzustellen, mit dem ein Blended Learning Konzept evaluiert werden kann. Blended Learning wird definiert als der kombinierte Einsatz von Präsenzlehre und computergestützter Lehre [Gr06]. Zu diesem Zweck stellen wir eine Lehrevaluation vor, die auf wesentlichen Kategorien der Lehrqualität [GS92], nämlich Lernerfolg und Akzeptanz aufbaut. Theorieleitet wird ein Modell der formativen Lehrevaluation entwickelt. Dieses stützt sich dabei auf die Erkenntnisse der Lerntaxonomie nach Krathwohl [Kr02, Kr64] zur Evaluation des Lernerfolgs und dem Technologieakzeptanzmodell (TAM) nach Davis [Da86]. Ergänzt werden diese zwei Konzepte der Lehrevaluation durch die Anwendung standardisierter Lehrevaluationen, die der Erhebung der Wahrnehmung des Lehrkonzepts dienen. Da es sich bei diesen um bereits erforschte Evaluationsmodelle [ZU07] handelt, stehen sie nicht im Fokus dieses Beitrags.

Im Folgenden werden die grundlegenden Theorien und Forschungsergebnisse vorgestellt und miteinander in Bezug gebracht. Es wird anhand einer Fallstudie das Modell exemplarisch angewandt. Anschließend stellen wir die Ergebnisse der Entwicklung der Evaluation vor und setzen uns im abschließenden Kapitel mit den Erkenntnissen, Grenzen und weiterführenden Forschungsergebnissen konstruktiv auseinander.

2 Definitionen

2.1 Lehrevaluationen von Blended Learning

Als Methode zur Bewertung und Analyse von Lehrveranstaltungen hat sich das Instrument der Lehrevaluation etabliert [ZU07]. Unter Evaluation wird die „systematische Analyse und empirische Untersuchung von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zum Zwecke ihrer Bewertung und Modifikation“ verstanden [Ri03]. Formative Ansätze zielen dabei auf die Verbesserung, während summative zur Wirkungsbeurteilung herangezogen werden [Kr01, Ri03]. Gegenstand der Lehrevaluation in Hinblick auf die formative Gestaltung der Lehre ist die Analyse der Lehrqualität. Der Begriff der Lehrqualität wird kontrovers diskutiert [SC06], nach Glowalla [GS92] bewegt sie sich in den Dimensionen der Lernzeit, des Lernerfolgs und der Akzeptanz des Mitteleinsatzes. Die Evaluation der Lernzeit [GS92] wurde in dem vorgestellten Evaluationsmodell nicht erhoben, da sie externen Faktoren wie der Definition im Zuge des Bologna-Prozesses [Va00] unterliegt. Im Weiteren wird die Evaluation des Lernerfolgs und der Akzeptanz der eingesetzten Technologie vorgestellt.

Oftmals werden notwendige Qualitätskriterien bei der Entwicklung einer Evaluation vernachlässigt [Ri03]. Sie spielen jedoch für die Nachvollziehbarkeit und Wiederverwendbarkeit der Ergebnisse eine bedeutende Rolle. In der Literatur werden verschiedene Qualitätskriterien diskutiert [GKV06]. Zweifelsfrei ist es vorteilhaft, Evaluationen in Hinblick auf statistische Gütekriterien und Kriterien der Handhabbarkeit zu prüfen. Kriterien der Reliabilität, Validität, Objektivität der Durchführung, der Auswertung und der Interpretation sowie Normierung, Ökonomie, Nützlichkeit und Vergleichbarkeit stellen dabei den wissenschaftlichen Rahmen [LiRa98, GKV06].

Zur Evaluation von Blended Learning wurden in der Vergangenheit verschiedene Verfahren angewandt. Beispielsweise werden im so genannten Mixed-Methods-Design qualitative und quantitative Fragebögen, Gruppeninterviews sowie studentische Prüfungsergebnisse zur Evaluation verwendet. Ziel dieser Studien ist oftmals die Verbesserung der Nutzerzufriedenheit der Studierenden oder die Entwicklung eines Evaluationswerkzeugs zur Einführung und Implementierung von Blended Learning im gesamten Studiengang [WO09, CSS17]. Mit quantitativen Befragungen, die mit der Analyse von Prüfungsergebnissen kombiniert werden, wird beispielsweise die Auswirkung zwischenmenschlicher Interaktion auf die Leistung und Zufriedenheit der Studierenden erforscht [CJ14]. Auch gibt es Forschung zur Nutzung qualitativer Methoden zur Evaluation von Blended Learning [HKW12]. Evaluationswerkzeuge fokussieren oft einzelne Aspekte einer Blended Learning Einführung wie die Erfahrung der Studierenden oder das Instruktionsdesign [AY08, HA14]. Das TAM wurde bereits für die Evaluation von Blended Learning bestätigt [BRP14]. Bislang liegt jedoch keine Studie vor, die auf die Evaluation mit dem Fokus auf Technologieakzeptanz von Blended Learning in Kombination mit einer Lernzielanalyse zielt.

2.2 Ermittlung der Lernziele

Grundlegendes Ziel der Lehre sollte das Erreichen von Lernzielen² sein. Um Lernziele beurteilen und beispielsweise nach ihrem Schwierigkeitsgrad organisieren zu können, haben sich als Referenz so genannte Lernzieltaxonomien etabliert. Die hier zitierte Lernzieltaxonomie [Kr02, Kr64] fußt auf den Ergebnissen von Bloom [BEFH56]. Sie ermöglicht die Bewertung der Lernzielerreichung. Das Modell enthält zwei verschiedene Dimensionen. Auf der x-Achse werden sechs kognitive Prozesse der Wissensadaption erfasst. Die kognitiven Prozesse werden mit den Verben „erinnern“, „verstehen“, „anwenden“, „analysieren“, „bewerten“ und „erzeugen“ beschrieben. Die Komplexität der Wissensadaption nimmt dabei zu. In dem Modell wird nicht zwangsläufig davon ausgegangen, dass die jeweils nächste (komplexere) Stufe die davorliegende enthält. Auf der y-Achse werden die Wissensdimensionen erfasst. Grundprinzip der Taxonomie ist es nun, dass Lernziele formuliert werden können, die sowohl den kognitiven Prozess als auch ein Thema enthalten und auf dieser Basis in das Modell eingeordnet werden

² Der Begriff Lernziel ist Klauer zufolge im Grunde unzulässig, da Lernen einen internen kognitiven Prozess bezeichnet, der nicht beobachtbar ist und für den demzufolge auch keine Ziele formuliert werden können. Richtig wäre es, von Lehrzielen zu sprechen [Klau87]. Der Ausdruck Lernziel ist jedoch in der Literatur geläufiger und wird daher in dieser Arbeit verwendet.

können. Es kann angenommen werden, dass das Erreichen höherer Lernziele zu einem höheren Lernerfolg führt [Kr02].

Die grundsätzliche Eignung einer Lernzieltaxonomie ist unumstritten. Zum einen lässt sie sich einsetzen, um der künftig auch im Deutschen Qualitätsrahmen (DQR) geforderten outcome-Orientierung, d.h. der Ausrichtung auf Lernergebnisse, gerecht zu werden. Zum anderen ermöglichen auf der Basis einer Taxonomie formulierte Lernziele eine präzise Überprüfung in Form kriteriumsorientierter Tests [KI87].

2.3 Erhebung der Technologieakzeptanz

Das TAM nach Davis [Da86] ist eines der meist zitierten Modelle im Bereich der Erforschung von Informationssystemen. Bis heute werden das Kernmodell und seine zahlreichen Erweiterungen verwendet und weiterentwickelt [KH06]. Das Ziel des TAM ist die Erklärung individuellen Verhaltens bezüglich einer Technologie [Da86]. Die Zielgröße des TAM ist die Operationalisierung der Benutzung einer Technologie. Dabei wirken die Faktoren wahrgenommene Nützlichkeit (Perceived Usefulness (PU)) sowie die wahrgenommene Einfachheit der Nutzung (Perceived Ease of Use (PEU)) auf die Einstellung (Attitude) des Nutzers, die wiederum die Benutzungsabsicht (Behavioral Intention) und damit die Nutzung beeinflusst. Das TAM wurde stark von verhaltenstheoretischen Ansätzen wie der Theory of Reasoned Action [FA75] geprägt.

Das TAM wird üblicherweise mit Hilfe von verschiedenen Items operationalisiert und seine Zusammenhänge werden in Form eines Fragebogens mit 7-Punkt-Likertskalen erhoben. Das hat den Vorteil, dass die Ergebnisse unterschiedlicher Studien miteinander verglichen werden können [KH06]. Auch wenn im Bereich der Akzeptanzmessung von Blended Learning Angeboten bisher nur wenige Studien zitiert werden können (vergl. exemplarisch [BRP14, TDP11, Yo16]), findet es auch in diesem Bereich seine Berechtigung. Schließlich verfügt das TAM über Vorteile, die es zu einem leicht handhabbaren Instrument machen, wie beispielsweise seine hohe Ergebnistransparenz durch standardisierte Items, seine leichte Anwendbarkeit, seine Unabhängigkeit von der Technologieumgebung und die wissenschaftliche Anerkennung der Maße. Häufig endet die Forschung im Bereich des TAM mit dem Erstellen des Modells, jedoch kann es auch genutzt werden, um die Gültigkeit der Prädiktoren im speziellen Anwendungsfall zu prüfen [FT12].

3 Beschreibung der Fallstudie

Die Fallstudie umfasst eine Lerngruppe von ca. 100 Studierenden der kooperierenden Studienprogramme „Master Wirtschaftsinformatik“ sowie „Master Betriebswirtschaftslehre“. Für die Vermittlung der basalen Lerninhalte werden Videos mit einer Länge von jeweils 10 bis 40 Minuten eingesetzt. Die Lernvideos ersetzen die klassische Präsenzvorlesung im Stil des Frontalunterrichts. Diese Tutorials beschreiben

auf der Basis einer Abfolge sprachlich erläuteter Folien grundlegendes Wissen zur Veranstaltung. Die zu Grunde liegenden Foliensätze werden den Studierenden zur Verfügung gestellt. In Form eines so genannten Reading Seminars werden die Studierenden an höher gelagerte Lernziele herangeführt und mit lehrstoffvertiefenden Sachtexten aus dem Forschungsbereich konfrontiert. Hierzu werden die Studierenden in Gruppen bis maximal 20 Personen eingeteilt, um Fragen zu den Texten zu bearbeiten und zu diskutieren. Für den Austausch der Texte, der Fragenkataloge und der Videos wird die digitale Lernplattform Stud.IP in Verbindung mit der Plattform „Open Cast“ zur Onlineverwaltung verwendet. Darüber hinaus werden den Studierenden ein veranstaltungsspezifisches Forum, die Möglichkeit der gemeinsamen Dateiverwaltung und wenn gewünscht auch ein Wiki zur Verfügung gestellt. Als erweiterte Form der Lehre werden zudem Vorträge von Partnern aus Industrie und Consulting organisiert.

4 Anwendung der Lehrevaluation

4.1 Vorgehen

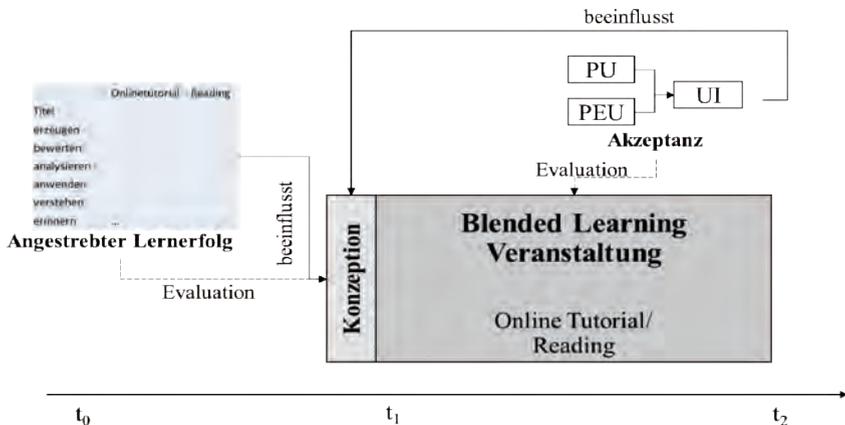


Abb. 1: Vorgehen der Evaluation

Die Evaluation des Blended Learning Konzeptes erfolgt wie in Abbildung 1 dargestellt in zwei Schritten zu unterschiedlichen Zeitpunkten. Der Lernerfolg wird im Zeitraum t_0 (Planung) erhoben. Dabei wird die inhaltliche Ausrichtung des Blended Learning Konzeptes evaluiert und ausgewertet. Die Ergebnisse können zum Beginn des Zeitraumes t_1 (Durchführung) noch Einfluss auf das Konzept des Blends nehmen. Die Akzeptanz der eingesetzten Medien wird also während der Durchführung der Veranstaltung erhoben (siehe Kapitel 4.2). Die Ergebnisse haben zum Auswertungszeitraum t_2 (Abschluss) formativen Einfluss auf die Konzeption in der Zukunft. Das Wechselspiel aus Analyse und formativem Eingriff zweier voneinander unabhängiger Evaluationsinstrumente ermöglicht die kontinuierliche Verbesserung der

Veranstaltung aus verschiedenen Blickwinkeln über mehrere Kohorten von Studierenden.

4.2 Evaluation des Lernerfolgs

Die Evaluation des Lernerfolgs [GS92] lässt sich mit Hilfe der Ermittlung der Lernzielerreichung operationalisieren. Zu diesem Zweck werden Lerninhalte definiert und mit den zu erwerbenden Kompetenzen verknüpft [BBF11]. Die Evaluation der Lernziele wird vorab durch die Lehrenden unter Einbezug der Lernzieltaxonomie nach Krathwohl vorgenommen [Kr02]. Vor dem Hintergrund der Entwicklung einer praktischen und leicht zu handhabenden Lösung wurde das Modell der Lernzieltaxonomie auf die Dimensionen der kognitiven Prozesse reduziert.

Ebene	Onlinetutorial	Reading
	Lerneinheit 8 Projektdurchführung	Lindner, F.; Wald, A.: Success factors of knowledge management in temporary organizations, in: International Journal of Project Management, 29 (2011), S. 877 – 888
erzeugen	-	-
bewerten	-	Sie können eine Einschätzung der hier vorgestellten Knowledge Management Methoden vornehmen und die Ergebnisse fallbezogen <i>interpretieren</i> und <i>bewerten</i> .
analysieren	Sie können die Earned Value Methode anwenden und die Ergebnisse <i>interpretieren</i>	Sie <i>verstehen</i> , welche Methoden wann verwendet werden und <i>können entscheiden</i> , wann die jeweiligen Methoden in anderen Situationen geeignet sind.
anwenden	Sie können Methoden der Qualitätssicherung im Projekt <i>anwenden</i> (Bsp. Earned Value)	Sie sind in der Lage, die im Text vorgestellten Methoden <i>anzuwenden</i> .
verstehen	Sie können den Begriff der Projektqualität <i>definieren</i>	Sie <i>verstehen</i> die aufgestellten Hypothesen und die Ergebnisse der Hypothesenprüfung.
erinnern	Sie erhalten einen Überblick über die Projektdurchführung	Sie können die Kernaussagen des Textes <i>wiedergeben</i> .

Tab. 1: Exemplarische Zuordnung der Lernziele der Fallstudie gemäß Krathwohl [Kr02, Kr64]

Der Inhalt der Vorlesung orientiert sich an anerkannten Zertifikaten [Pr11, Pr13]. Die grundsätzliche inhaltliche Qualität kann somit als gesichert angesehen werden.

Anschließend werden für die zuvor definierten Lerneinheiten die Lernziele formuliert und eingeordnet. Tabelle 1 zeigt dies am Fallbeispiel. Die auf diese Weise evaluierten Lernziele dienen als Basis für die Formulierung des Leistungsanspruchs der Veranstaltung. Sie können in Form von Prüfungen (schriftliche Klausuren) anschließend leicht geprüft werden.

4.3 Evaluation der Technologie-Akzeptanz

Die Akzeptanz [GS92] als notwendiges Maß zur Beschreibung der Lehrqualität wird mit Hilfe des TAM gemessen. Auch wenn die einheitliche Handhabung des TAM dieses zum Standardmodell in der Akzeptanzmessung gemacht hat, ist bei der Erhebung der Akzeptanz stets zu berücksichtigen, dass sich die Anwendungssituationen unterscheiden. Nicht jede studentische Kohorte gleicht der vorhergehenden Generation, der wahrgenommene Leistungsdruck variiert aufgrund von Faktoren, die außerhalb des Einflussbereichs des Lehrenden stehen. Daher war es bei der Konstruktion des Modells wichtig, das Modell leicht anwendbar und allgemein genug zu gestalten, um eine hohe Transparenz zu gewährleisten. Zudem wurden bei der Konstruktion des zu prüfenden Modells einige Vorbedingungen formuliert (Tabelle 2), die auch bei der Bewertung des gesamten Evaluationsmodells hinsichtlich wissenschaftlicher Kriterien [GKV06] (vergl. Kapitel 5) von großer Bedeutung sind.

Bedingung	Verweis	Umsetzung
Die Akzeptanzmaße müssen geprüft bzw. anerkannt und standardisiert sein	[KH06, VE03]	Standardmaße aus dem Original-TAM, bzw. Verwendung bereits geprüfter Operationalisierungen
Geringe Anzahl an Faktoren, zur statistischen Überprüfbarkeit auch bei geringer Studierendenzahl	[Bo06]	Reduktion auf wenige, wohl bewiesene Maße
Unabhängigkeit der Maße	[GKV06]	Überprüfung der Reliabilität
Akzeptanzmaße sollten vom Lehrstoff unabhängig sein	[Vo13]	Verzicht auf Maße wie spezifische Erfahrung bzw. Maße für bestimmte Arten von Technologien (z.B. Perceived Playfulness für Spiele) die kontextspezifisch sind

Tab. 2: Anforderungen an das Modell zur Erhebung der Akzeptanz von Blended Learning

Zur Evaluation unseres Konzeptes wurde ein Modell im Sinne des TAM von Davis gebildet. Die TAM-Konstrukte Perceived Ease of Use und Perceived Usefulness [BH03] bilden die Kernkonstrukte des Modells. Tabelle 3 zeigt die Operationalisierung der Konstrukte. Die Bewertung durch die Studierenden erfolgte im laufenden Semester. Die Untersuchung der Reliabilität zeigt, dass sich die standardisierten Items grundsätzlich eignen, um die Konstrukte zu erheben (0,822 (PEU) und 0,982 (PU)). Die Regressionsanalyse zeigt, dass nahezu 60% der Use Intention mit Hilfe des Modells erklärt werden können (adjusted $R^2=0,595$). Dies ist ein verhältnismäßig hohes Maß

[KH06]. Hauptprädiktor der Use Intention ist die PU ($\beta = 0,739$), die damit etwa die Hälfte der erklärten Use Intention beeinflusst (vergl. auch [Yo16]). Es wurden auch soziale Faktoren wie Subjective Norm nach Ajzen ([AF80]) und das Konstrukt der Perceived Enjoyment nach Davis [Da86] und van der Heijden [He04] geprüft. Diese zeigen kaum signifikanten Einfluss auf die Nutzungsabsicht, was auf die hohe Aufgabenorientierung der Fallstudie zurückzuführen ist. Daher wurden diese Prädiktoren nicht in das Modell übernommen.

Konstrukt	Definition	Item
PU	“the degree to which a person believes that using a particular system would enhance his or her job performance” [Da86]	Die Nutzung der Videos verbessert meine Lernumgebung. ... steigert meine Produktivität beim Lernen. ... fördert meine Effizienz beim Lernen. Ich empfinde das System nützlich für meinen Lernprozess.
PEU	“the degree to which a person believes, that using a system would be free of effort” [Da86]	Mit den Videos zu arbeiten erfordert nicht viel mentale Anstrengung. Ich empfinde die Videos als einfach zu nutzen. Ich finde es einfach mit den Videos so zu arbeiten wie ich es möchte.

Tab. 3: Konstruktion der Fragebogen Items zur Erhebung der Technologieakzeptanz

Ergänzend zu den Konstrukten des TAM wurden die Studierenden gebeten, ihr persönliches Empfinden bezüglich der Gestaltung des Lernalltags in dieser Blended Learning Veranstaltung auszudrücken. Mit Hilfe einer Binominalskala (trifft zu/ trifft nicht zu) [Bo06, ZU07] wurden folgende Aussagen bewertet: *Ich erhalte einen guten Überblick über das Thema; Ich arbeite kontinuierlich über das Semester mit; Ich fühle mich gut auf die Prüfung vorbereitet; Ich kann den Lehrstoff gut bewältigen; Ich kann selbständig arbeiten; Ich kann mein Arbeitstempo selbst bestimmen.*

5 Analyse des Evaluationsverfahrens

Die hier vorgestellte Evaluationsmethodik ist im Bereich der Selbstevaluation [BBF11] mit dem Zweck der interaktiven Evaluation zu verorten. Die Tabelle 4 gibt Aufschluss darüber, wie sich die Maße interpretieren lassen.

In Anlehnung an Tabelle 4 sind das Fehlen höherer Lernzielebenen bzw. das Stagnieren auf unteren Lernzielebenen Anzeichen für dringenden Handlungsbedarf, die Konzeption zu überarbeiten. Ebenso sollten Abweichungen der statistischen Werte, wie nicht signifikante Werte für UI, PU und PEU in Maßnahmen zur Verbesserung der Lehre münden. Bei geringer Signifikanz der wahrgenommenen Nützlichkeit müssen

beispielsweise die Inhalte des Blends sowie die Verfahren zur Überprüfung des Lernfortschritts (in den meisten Fällen eine schriftliche Prüfung) hinsichtlich ihrer Übereinstimmung geprüft werden. Aufschluss darüber kann mit Hilfe der formulierten Lernziele erlangt werden.

t	Was	Wie	Interpretation
t_0	Qualität/ Lernerfolg	Orientierung an Lerninhalten anderer vergleichbarer Institutionen	Besteht eine weitgehende Übereinstimmung inhaltlich und ermöglicht der Blend das Erlangen inhaltlich qualitativen Lernstoffs, ist dies vorteilhaft.
	Lernerfolg/ Lernniveau	Zuordnung und Vergleich anhand von Taxonomien	Werden grundlegende Lernlevel erreicht, erhöht der Blend die Lernzielerreichung, ist dies als Lernerfolg zu bewerten.
t_1	Akzeptanz	TAM	Hohe Werte von UI und den Prädiktoren lassen auf eine hohe Akzeptanz schließen.
	Selbstorga- nisation	Binominalskala	Werden die Rahmenbedingungen zur eigenständigen Lernorganisation als gegeben angesehen, ist dies positiv zu bewerten.

Tab. 4: Inhaltliche Auswertung der Evaluation

Das hier vorgestellte Verfahren kann einige Kritikpunkte und Anforderungen [GKV06, ZU07] an Lehrevaluationen entkräften bzw. erfüllen. Die Konstruktion erfolgt theoriegeleitet (Krathwohl und TAM) auf Basis anerkannter -wenn auch kontextfremder- Vorgehensweisen und kann daher den Ansprüchen der Reliabilität, Validität, Objektivität und Vergleichbarkeit Rechnung tragen. Die Gefahr der subjektiven Interpretation der Ergebnisse [BBF11] durch den Lehrkörper besteht im Bereich der Akzeptanzevaluation nicht, im Bereich des Lernerfolgs nur bedingt. Hier empfiehlt sich eine zusätzliche Kontrollinstanz bspw. mittels Sichtung durch weitere Dozenten. Das Verfahren ist demnach als reliabel einzuschätzen. Es werden Maße verwendet, die in anderen Kontexten bereits geprüft wurden und erfolgreich zur Anwendung kamen, damit lässt sich eine erwünschte Normierung vornehmen. Die Evaluation des Lernerfolgs und die Formulierung der Lernziele ist mit einem sehr hohen Initialaufwand verbunden. Die gut dokumentierte Darstellung der Lernlevel ist jedoch äußerst nützlich. Sie vereinfacht die Ableitung potenziellen Handlungsbedarfs massiv. Die Evaluation der Akzeptanz ist mit Hilfe standardisierter Konstrukte einfach und daher ökonomisch. Beide Verfahren lassen, wie bereits beschrieben, die Ableitung von Handlungsempfehlungen zu und unterstützen sich gegenseitig in ihren Aussagen in diesem Prozess. Der Umfang der Evaluation ist in weiten Teilen handhabbar und nicht zu hoch [ZU07]. Damit erfüllt die hier vorgestellte Evaluationsvariante wichtige Qualitätskriterien. Aufgrund des hohen Bezugs zur Technologiekzeptanz im Zusammenspiel mit den Inhalten der betrachteten Lehrveranstaltung schließt das vorgestellte Verfahren die zuvor formulierte Forschungslücke und eignet sich auch dazu, bestehende Verfahren durch neue Blickwinkel zu ergänzen.

6 Grenzen und Ausblick

Das vorgestellte Modell dient zur Messung der Eignung eines Blended Learning Ansatzes. Wichtig ist zu betonen, dass hier nicht die Leistung des Lehrenden im Zentrum steht, sondern die Integration der Veranstaltung in den Lernalltag der Studierenden. Zur Erhebung der Qualitäten des Lehrenden existieren bereits zahlreiche Standardverfahren, die ergänzend zur hier vorgestellten Variante stattfinden. Erweiterungen der Lernumgebung wie beispielsweise die vielbeschriebene Gamification können ebenso in das Evaluationsmodell des TAM übernommen werden, wie Quizmodule oder E-Tests. Hierfür existieren zahlreiche weitere Konstrukte im TAM. Zusätzlich lassen sich auch Strukturen aus Lernkompetenzmodellen mit dieser Art der Evaluation verknüpfen. Dabei ist ein besonderer Forschungsfokus auf den langfristigen Einsatz der Evaluation mit Hilfe des Fragebogens zu legen, um Ergebnisse zu vergleichen, Effekte von Maßnahmen zu analysieren und konkrete Handlungsempfehlungen und deren Auswirkungen empirisch belegbar zu machen.

Literaturverzeichnis

- [AF80] Ajzen, I.; Fishbein, M.: Understanding attitudes and prediction social behavior. Prentice Hall, Englewood Cliffs NJ, 1980.
- [AY08] Akkoyunlu, B.; Yılmaz-Soylu, M.: Development of a scale on learners' views on blended learning and its implementation process. In: Internet High. Educ. 11/08, Nr. 1, S. 26–32, 2008.
- [BRP14] Bachtiar, F.A.; Rachmadi, A.; Pradana, F.: Acceptance in the deployment of blended learning as a learning resource in information technology and computer science program. In: Computer Aided System Engineering (APCASE), Asia-Pacific Conference on: IEEE, S. 131–135, 2014.
- [BBF11] Beywl, W.; Bestvater, H.; Friedrich, V.: Selbstevaluation in der Lehre: ein Wegweiser für sichtbares Lernen und besseres Lehren. Waxmann, Münster, 2011.
- [BEFH56] Bloom, Benjamin S.; Englehart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R.: Taxonomy of educational objectives-handbook 1: Cognitive domain. David McKay Company Inc, New York, 1956.
- [BH03] Benbunan-Fich, R.; Hiltz, S.R.: Mediators of the effectiveness of online courses. In: IEEE Transactions on Professional Communication (46/2003), Nr. 4, S. 298–312.
- [Bl00] Bligh, D. A.: What's the use of lectures. Jossey-Bass Publishers, San Francisco, 2000.
- [Bo06] Bortz, J.: Statistik für Human und Sozialwissenschaftler. 6. Aufl., Springer-Verlag, Berlin, 2006.
- [CJ14] Chan D.; Jia W.: The Effect of Human Interactions on Student Performance and Satisfaction of Blended Learning. In: Academy of Educational Leadership Journal

18/14, Nr. 3, S. 11–21, 2014.

- [CSS17] Chmiel, A.S.; Shaha, M.; Schneider, D.K.: Introduction of blended learning in a master program: Developing an integrative mixed method evaluation framework. In: *Nurse Education Today* 48/17, S. 172–179, 2017.
- [Da86] Davis, F.D.: *A technology acceptance model for empirically testing new end-user information systems: Theory and results*, Massachusetts Institute of Technology, 1986.
- [FA75] Fishbein, M.; Ajzen, I.: *Belief, Attitude, Intention and Behavior – An Introduction to Theory and Research*. Addison-Wesley, Reading MA, 1975.
- [FT12] Freundlieb, M.; Teuteberg, F.: Augmentierte Nachhaltigkeitsberichterstattung. In: *HMD Praxis der Wirtschaftsinformatik* 49/12, Nr. 4, S. 104–112, 2012.
- [GS92] Glowalla, U.; Schoop, E.: *Hypertext und Multimedia: Neue Wege in der computerunterstützten Aus- und Weiterbildung* GI-Symposium Schloß Rauschholzhausen Tagungsstätte der Universität Gießen 28.-30.4.1992.
- [GKV06] Gollwitzer, M.; Kranz, D.; Vogel, E.: Die Validität studentischer Lehrveranstaltungsevaluationen und ihre Nützlichkeit für die Verbesserung der Hochschullehre: Neuere Befunde zu den Gütekriterien des „Trierer Inventars zur Lehrevaluation“ (TRIL). In: *Didaktik und Evaluation in der Psychologie*, S. 90–104, 2006.
- [Gr06] Graham, C.R.: Blended learning systems. In (Bonk, C.J.; Graham, C.R. Hrsg.): *The Handbook of Blended Learning*, Pfeiffer, San Francisco, S. 3–21, 2006.
- [HKW12] Harding, A.; Kaczynski, D.; Wood, L.: Evaluation of blended learning: analysis of qualitative data. In: *Proceedings of The Australian Conference on Science and Mathematics Education (formerly UniServe Science Conference)*. Bd. 11, 2012.
- [He04] Heijden, H. van der: User Acceptance of Hedonic Information Systems. In: *MIS Quarterly* 28/04, Nr. 4, S. 695–704, 2004.
- [HA14] Halverson, L.R.; Graham, C.R.; Spring, K.J.; Drysdale, J.S.; Henrie, C.R.: A thematic analysis of the most highly cited scholarship in the first decade of blended learning research. In: *Internet High. Educ.* 20/14, S. 20–34, 2014.
- [KH06] King, W.R.; He, J.: A meta-analysis of the technology acceptance model. In: *Information & Management* 43/06, Nr. 6, S. 740–755, 2006.
- [K187] Klauer, K.J.: *Kriteriumsorientierte Tests*. Hogrefe, Göttingen, 1987.
- [Kr02] Krathwohl, D.R.: A revision of Bloom’s taxonomy: An overview. In: *Theory into practice* 41/02, Nr. 4, S. 212–218, 2002.
- [Kr64] Krathwohl, D.R.: *Taxonomy of educational objectives: The classification of educational goals*. Bd. 2, Longmans, Green, New York, 1964.
- [Kr01] Kromrey, H.: Evaluation von Lehre und Studium - Anforderungen an Methodik und Design. In (Spiel, C., Hrsg.): *Evaluation universitärer Lehre – Zwischen Qualitätsmanagement und Selbstzweck*. Waxmann, Münster, S. 21–60, 2001.
- [LiRa98] Lienert, G.A.; Raatz, U.: *Testaufbau und Testanalyse*. Beltz Weinheim 1998.

- [Pr11] Projektmanagement, Deutsche Gesellschaft für: Projektmanagement-Fachmann. 10. Aufl., Bd. 1, Wissenschaft & Praxis, Sternenfels, 2011.
- [Pr13] Project Management Institute (Hrsg.): A Guide to the Project Management Body of Knowledge (PMBOK®Guide)-Fifth Edition. Project Management Institute, Pennsylvania, 2013
- [Ri03] Rindermann, H.: Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. In: Zeitschrift für Evaluation 02/03, S. 233–256, 2003.
- [Ru14] Rummler, M. (Hrsg.): Vorlesungen innovativ gestalten: neue Lernformen für große Lerngruppen; Lehren an der Hochschule. Beltz, Weinheim, 2014.
- [SC06] Schulz, N.; Greve, W.; Koch, U.; Koops, T.; Wilmers, N.: Wie gut erfassen Fragebögen die Qualität der Lehre. In (Krampen, G.; Zayer, H. Hrsg.): Psychologiedidaktik und Evaluation. Hofgreffe, Göttingen, S. 75–89, 2006.
- [SSR16] Staufenbiel, T.; Seppelfricke, T.; Rickers, J.: Prädiktoren studentischer Lehrveranstaltungsevaluationen: Eine Mehrebenenanalyse. In: Diagnostica 62/16, Nr. 1, S. 44–59, 2016.
- [TDP11] Tselios, N.K. ; Daskalakis, S.; Papadopoulou, M.: Assessing the Acceptance of a Blended Learning University Course. In: Educational Technology & Society 14/11, Nr. 2, S. 224–235, 2011.
- [Va00] Van der Wende, M.C.: The Bologna Declaration: Enhancing the transparency and competitiveness of European higher education. In: Journal of Studies in International Education Bd. 4 (2000), Nr. 2, S. 3–10.
- [VE03] Venkatesh, V.; Morris, M.G.; Davis, G.B.; Davis, F.D.: User Acceptance of Information Technology: Toward a Unified View. In: MIS Quarterly 27/03, Nr. 3, S. 425–478. 2003.
- [Vo13] Vogelsang, K.: Analyse der Softwareakzeptanz bei der Anwendung von Projektmanagement-Software und Ableitung von Handlungsempfehlungen (Schriftenreihe Innovative Betriebswirtschaftliche Forschung und Praxis). Verlag Dr. Kovac, Hamburg, 2013.
- [WO09] Woltering, V.; Herrler, A.; Spreckelsen, C.; Spitzer, K.: Blended learning positively affects students' satisfaction and the role of the tutor in the problem-based learning process: results of a mixed-method evaluation. In: Advances in Health Sciences Education, S. 725–738, 2009.
- [Yo16] Yoshida, H.: Perceived Usefulness of „Flipped Learning“ on Instructional Design for Elementary and Secondary Education: With Focus on Pre-service Teacher Education. In: International Journal of Information and Education Technology 06/16, Nr. 6, S. 430, 2016.
- [ZU07] Zumbach, J.; Spinath, B.; Schahn, J.; Friedrich, M.; Kögel, M.; others: Entwicklung einer Kurzsкала zur Lehrevaluation. In: Psychologiedidaktik und Evaluation VI. Vandenhoeck & Ruprecht unipress, Göttingen, S. 317–325, 2007.

Schnell zu erfassen: Ein Komplexitätsmaß für Mehrfachauswahlfragen

Frauke Kämmerer¹, Heinrich Söbke² und Lukas Hartung²

Abstract: Mehrfachauswahlfragen können für Lern- und Prüfungszwecke eingesetzt werden. Bei der Nutzung zu Unterhaltungszwecken – wie in Quiz-Apps – ist eine um ein Vielfaches größere Popularität zu beobachten. Eine Analyse möglicher Ursachen für die hohe Attraktivität lässt u.a. eine geringere textuelle Komplexität der Fragen in Unterhaltungskontexten vermuten. Auf Grundlage existierender textueller Komplexitätsmaße haben wir daher eine Metrik für die textuelle Komplexität einer Mehrfachauswahlfrage entwickelt. Zur Validierung wurde die entstandene Metrik analytisch auf zwei Corpora von Mehrfachauswahlfragen angewendet. Die Fragen des Corpus mit Fachfragen zeigten sich deutlich komplexer als die Fragen des Unterhaltungs-Corpus. Darauf aufbauend wurden die Fachfragen so überarbeitet, dass sie entsprechend der Metrik eine geringere Komplexität zeigten. Alle drei Gruppen von Fragen wurden in einer Probanden-Studie (n=14) wettbewerbsorientiert in einer Quiz-App beantwortet. Obwohl eine Annäherung erzielt werden konnte, wiesen die Unterhaltungsfragen weiterhin bessere Werte bzgl. des Spielspaßes auf. Wesentliche Ergebnisse der Arbeit sind ein Entwurf einer Komplexitätsmetrik für Mehrfachauswahlfragen, deren Validierung und studiengestützte Hinweise, dass eine konstruktive Nutzung der Metrik zu leichter lesbaren Fragen führt.

Keywords: Multiple Choice Question; Metrik; Quiz; Instructional Design; Lesbarkeitsindex

1 Einleitung

Mehrfachauswahlfragen (Multiple-Choice-Fragen, MCQ) können z. B. zur Überprüfung des Lernfortschritts eingesetzt werden. Gleichzeitig lassen sie sich auch zum Lernen nutzen, denn durch wiederholtes Beantworten von Fragen werden Lernprozesse initiiert („Testing-effect“, [BR07]). Eine derartige Fragenwiederholung lässt sich auch in kommerziellen Quiz-Apps beobachten [Ha14], die z. T. hohe Benutzerzahlen aufweisen [Ru14]. In früheren Arbeiten haben wir beobachtet, dass Spieler von Quiz-Apps diese mit einer Lernerwartung und -bereitschaft nutzen [Sö15]. Gleichzeitig wird ein teilweise vorhandener Zeitdruck, beispielsweise ein Zeitlimit von zehn Sekunden in der App *QuizUp* [P114], bei der Beantwortung der Fragen als eher abträglich für den Lernerfolg gesehen, da Fachfragen tendenziell länger und insgesamt schwerer erfassbar zu sein scheinen als Fragen der Unterhaltungsthemen [SW16]. Tabelle 1 zeigt jeweils ein Beispiel einer Unterhaltungsfrage des Themengebietes *Allgemeinwissen* und eine Fachfrage des Themas *Technische Luftreinhaltung*. Es ist anzunehmen, dass die

¹ Hochschule Nordhausen, Weinberghof 4, 99734 Nordhausen, frauke.kaemmerer@fh-nordhausen.de

² Bauhaus-Universität Weimar, Bauhaus-Institut für zukunftsfähige Infrastruktursysteme, Coudraystr. 7, 99423 Weimar, {heinrich.sobke|lukas.hartung}@uni-weimar.de

unterschiedliche textuelle Komplexität nur eine mögliche Ursache für das Zurückbleiben der Akzeptanz von kommerziellen Quiz-Apps in Lernkontexten hinter deren Popularität als Unterhaltungsmedium ist. In multikausalen Erklärungsmodellen spielen sicherlich auch weitere Gründe wie unterschiedliche Nutzerpräferenzen, der Lernkontext selbst sowie die Beschränkung auf festgelegte Themen eine Rolle. Jedoch lässt sich für die *Textuelle Komplexität* mit Metriken in Form von objektiv feststellbaren Kennzahlen analytisch (bei existierenden Fragen) und konstruktiv (beim Fragenentwurf) automatisiert mit relativ geringem Aufwand Abhilfe schaffen [SK16].

<u>Thema:</u> Allgemeinwissen	<u>Thema:</u> Technische Luftreinhaltung
Aus welcher Bohnenart wird Tofu hergestellt?	Welche Geschwindigkeit wird verwendet, um den Widerstandsbeiwert bei Biofiltern zu berechnen?
<u>A: Sojabohnen</u>	<u>A: Anströmungsgeschwindigkeit</u>
B: Weiße Bohnen	B: Durchschnittsgeschwindigkeit
C: Grüne Bohnen	C: Ausströmungsgeschwindigkeit
D: Kidneybohnen	D: Schallgeschwindigkeit

Tabelle 1 Beispiel einer Unterhaltungs- und einer Fachfrage

2 Textuelle Komplexität und Komplexitätskriterien

2.1 Grundlagen

Metriken zur Messung des Leseaufwandes von Auswahlfragen scheinen bisher nicht im Fokus der Lehr-Lern-Forschung gestanden zu haben. So finden sich zwar Richtlinien für den Fragenentwurf (z.B. [HR13]), nicht aber konkrete Kennzahlen. Hingegen gibt es für Texte im Allgemeinen mehrere Lösungsansätze. Bamberger und Vanecek [BV84] beschreiben Lesbarkeitsformeln, die formale Textmerkmale in eine Maßzahl für die Verständlichkeit abbilden. Um zu einem ersten Entwurf einer Metrik zu gelangen, wird hier vereinfachend davon ausgegangen, dass Leseaufwand und Verständlichkeit miteinander korrelieren. Oft kann eine Metrik nicht ohne Anpassung auf eine andere Sprache übertragen werden, da Textmerkmale genutzt werden, die von Sprache zu Sprache divergieren.

In der Softwareentwicklung ist es gleichfalls essentiell, Quelltexte möglichst schnell und korrekt erfassen zu können. Da es sich hier um formal strukturierte Texte handelt, sind Metriken für diese besonders leicht zu entwerfen und anzuwenden.

Sowohl für die Lesbarkeit von Texten als auch für die Komplexität von Quellcode gibt es eine Vielzahl von Metriken. Daher haben wir uns im Rahmen dieser Arbeit auf fünf Verfahren beschränkt, die sich im Rahmen der Literaturrecherche als relevant herausstellten. Diese wurden bezüglich Eingangsgrößen, Rechenvorschriften und Ausgangswert analysiert (s. Tabelle 2).

Metrik	Beschreibung	Wesentliche Eingangsgrößen
Flesch-Reading-Ease-Formel (FRE)	Maßzahl zur Berechnung der Lesbarkeit eines Textes [Fl48], deutsche Bearbeitung [Am78]	<ul style="list-style-type: none"> • Durchschnittliche Satzlänge • Durchschnittliche Silbenanzahl pro Wort
Gunning-Fog-Index (GFI)	Maßzahl zur Beschreibung der Lesbarkeit eines Textes [Gu52]	<ul style="list-style-type: none"> • Anzahl der Wörter • Anzahl der Sätze • Anzahl der Wörter mit mehr als zwei Silben
Halstead-Metriken	Verfahren zur Komplexitätsmessung von Softwarequelltexten, Beispiel: Kennzahl D : Schwierigkeit, ein Programm zu verstehen [Ha77]	<ul style="list-style-type: none"> • Jeweils Anzahl unterschiedlicher Operanden und Operatoren • Jeweils Gesamtanzahl von Operanden und Operatoren
Wiener Sachtextformel (WSTF)	Maßzahl zur Berechnung der Lesbarkeit von deutschsprachigen Texten [BV84]	<ul style="list-style-type: none"> • Anteil der Wörter mit mehr als zwei Silben • mittlere Satzlänge • Anteil der Wörter mit mehr als sechs Buchstaben • Anteil der einsilbigen Wörter
Zyklomatische Komplexität (ZK)	Messung der Komplexität von Software, um Verständnis-kritische Stellen zu identifizieren, auch McCabe-Metrik genannt [Mc76]	<ul style="list-style-type: none"> • Anzahl bedingter Anweisungen • Anzahl der Unterfunktionen

Tabelle 2 Untersuchte Metriken und ihre Eingangsgrößen

2.2 Das Komplexitätsmaß m_{Kt}

Die oben beschriebenen Metriken wurden für längere Texte entwickelt. Eine Auswahlfrage besteht aber in der Regel nur aus etwa ein bis zwei Sätzen. Um das

Verhalten der Metriken auch bei Anwendung auf derartige Beispiele nachvollziehen zu können, wurden zunächst die Berechnungsvorschriften analysiert und eine Liste der möglichen Eingangsgrößen erstellt. Zusätzlich wurden für ca. 20 ausgewählte Beispielfragen mit jeweils einem besonders markant ausgeprägten Merkmal (z. B. viele Fremdwörter, kurze Fragen, lange Antwortoptionen) die zugehörigen Metrikwerte bestimmt. Die verschiedenen Berechnungsvorschriften ergeben auf eine Frage angewendet unterschiedliche Werte für die Verständlichkeit bzw. Lesbarkeit. So hat FRE einen typischen Wertebereich von 0-100, wobei 0 als sehr schwer gilt. Der Wert des GFI hingegen entspricht der jeweiligen Schulklassenstufe, in der diese Fertigkeit erreicht sein sollte. Um die verschiedenen Ergebnisse vergleichbar zu machen, wurden mit Hilfe einer Tabelle die verschiedenen Wertebereiche auf Klassen von 1 bis 6 abgebildet („Normierung“, s. Tabelle 3) Daraus konnte im Wesentlichen die Bedeutung der jeweiligen Eingangsgrößen abgeleitet werden. Mit diesem Wissen und dem Ziel einer möglichst einfachen und automatisierbaren Metrik wurde dann mit verschiedenen Varianten experimentiert.

		Metrik				
		Halstead	ZK	WSTF	FRE	GFI
Normierung	1	bis 1,4	bis 3,5	bis 5,1	über 65,6	bis 2,11
	2	1,5 - 2,4	3,6 – 5,0	5,2 – 8,4	65,6 – 43,9	2,12 – 3,11
	3	2,5 - 3,4	5,1 – 6,5	8,5 – 11,7	43,8 – 22,2	3,12 – 4,11
	4	3,5 - 4,4	6,6 – 8,0	11,8 – 15,0	22,1 – 0,5	4,12 – 5,11
	5	4,5 - 5,4	8,1 – 9,5	15,1 – 18,3	0,4 – -21,3	5,12 – 6,11
	6	ab 5,5	ab 9,6	ab 18,4	unter -21,3	ab 6,12

Tabelle 3 Normierungstabelle der Metriken

Abbildung 1 zeigt eine Beispielfrage, die sehr hohe Werte bezüglich des Merkmals Fach- und Fremdwörter ($x=3$ bei einem Mittelwert von 1,4) und langer Wörter ($x=13$, Mittelwert: 7,0) hat. Die normierte Bewertung einer solchen Frage ist in Tabelle 4 zu sehen.

Zu welchem umweltschädlichen Phänomen kann die Emission von Schwefeldioxid (z.B. bei industriellen Verbrennungsprozessen) führen?	
Saurer Regen	Verstärkung Treibhauseffektes
Extreme Wolkenbildung	keine besondere Auswirkung

Abbildung 1: Beispielfrage mit vielen langen Worten und Fremdworten

Metrik	Halstead	ZK	WSTF	FRE	GFI
Wert	16	7	13,85	7,53	6,64
Normierter Wert	6	4	4	4	6

Tabelle 4: Werte und normierte Werte der Metriken für die Frage aus Abbildung 1

Die entworfene Metrik m_{Kt} enthält sieben Eingangsgrößen. Zu diesen weitestgehend objektiv feststellbaren Indikatoren zählen die Wortanzahl im Fragenstamm (W_F) und in den Antwortoptionen (W_A), die Anzahl der langen Wörter (W_L), der Fremd- oder Fachwörter (W_{Fr}) und der Satzzeichen (SZ) sowie eine gleichmäßige Länge der Antwortoptionen (GL_A) und die Nicht-Nutzung von Signalwörtern in den Antwortoptionen (SW). Signalwörter sind Wörter wie „immer“ oder „nie“, die einer Antwort einen absoluten Anspruch geben, der jedoch gewöhnlich nicht erfüllbar ist und damit diese Antwort als richtige ausschließen [Po93].

Indikatoren, die nur mit Hilfe von Fachwissen bestimmbar sind, d.h. nicht oder nur sehr schwer automatisiert bestimmbar sind, wurden nicht weiter betrachtet. Darunter fallen u.a. die Bewertungen, ob es sich um Fang- oder Trickfragen handelt, wie sinnhaft die Distraktoren sind und ob die Frage auch ohne Vorgabe von Antwortmöglichkeiten zu beantworten ist. Die ausgewählten Indikatoren wurden in einer Berechnungsvorschrift zu einer Kennzahl, dem Maß für die textuelle Komplexität m_{Kt} miteinander verknüpft. Dabei wurde jedem Indikator gemäß einer Transformationstabelle ein ganzzahliger Wert zwischen 0 bis 3 zugewiesen. Tabelle 5 zeigt als Beispiel die Transformationstabelle für W_F . Eine Ausnahme ist der Indikator SW. Hier erfolgte eine dichotome Bewertung: 0 für die Abwesenheit von Signalwörtern, 3 falls welche vorhanden waren. Als langes Wort gilt eines ab drei Silben. Die Festlegung als Fremd- oder Fachwort kann über das Vorhandensein in einem (Fach-)Wörterbuch erfolgen, unterliegt jedoch auch subjektiven Einflüssen.

Anzahl Wörter	Wert W_F
≥ 13	3
10-12	2
7-9	1
≤ 6	0

Tabelle 5 Transformationstabelle Indikator *Wortanzahl im Fragenstamm* (W_F)

Formel 1 ist die hier genutzte Berechnungsvorschrift zur Bestimmung eines Maßes für die textuelle Komplexität m_{Kt} . Durch die den Wertebereich der sieben Summanden von 0 bis 3 kann der Wert dieser Metrik zwischen 0 und 21 schwanken. Dabei zeigen geringere Werte auch eine geringere Komplexität an.

$$m_{Kt} = W_F + W_A + W_L + W_{Fr} + SZ + GL_A + SW \quad (1)$$

2.3 Validierung

Diese Berechnungsvorschrift ist nicht als allgemein gültig zu sehen, sondern wurde im Rahmen dieser Betrachtungen entwickelt und hat sich für Beispielfragen als funktionierend erwiesen. Es sind jedoch auch weitere sinnvolle Alternativen denkbar, u.a. mit stetigen Transformationsfunktionen sowie zusätzliche Gewichtungen der einzelnen Summanden.

Zur Abklärung der Funktionalität der Berechnungsvorschrift wurde sie auf weitere Fragen angewendet. Zum einen waren dies Fragen zum Thema *Allgemeinwissen*. (vgl. Tabelle 1). Diese wurden aus dem beliebten entsprechenden Topic *Allgemeinwissen* der Quiz-App *QuizUp* extrahiert und sind daher als typische Unterhaltungsfragen zu bewerten. Ein weiterer Fragencorpus hatte das fachliche Thema *Anaerobtechnik*. Tabelle 6 zeigt die Ergebnisse der Metrik für diese Corpora. Es ist ersichtlich, dass die Metrik die vermutete unterschiedliche Komplexität von Fach- und Unterhaltungsfragen widerspiegelt. Mit 7,2 bzw. 7,8 ist der Wert für Fachfragen höher als für die Unterhaltungsfragen (4,8). Wesentliche Gründe hierfür sind die höhere Anzahl an Fremd- und Fachwörtern, langen Wörtern und längere Antwortoptionen. Bezüglich der gleichen Länge der Antwortoptionen besteht noch Optimierungspotenzial bei den

Fachfragen. Signalwörter sind hingegen kaum vorhanden. Hier scheinen die Richtlinien zur Erstellung von Auswahlfragen bereits in der Praxis angekommen zu sein.

Attribute/ Indikatoren	Corpus	Allgemein- wissen	Technische Luftrein- haltung	Anaerob- technik (orig.)	Anaerob- technik (rev.)
QuizUp-Topic		<i>Lukas‘ Allgemeinwis- sen</i>	<i>Luftikusse</i>	<i>Anaerobier Stout</i>	<i>Anaerobier Lager</i>
Anzahl der Fragen		109	162	40	40
W_F		1,4	1,5	1,3	1,0
W_A		0,7	1,1	1,3	0,7
W_L		1,1	2,3	2,6	2,1
W_{Fr}		0,4	1,1	1,6	1,1
SZ		1,1	0,9	0,5	0,3
GL_A		0,2	0,3	0,4	0,2
SW		0,0	0,04	0,2	0,0
m_{Kt}		4,8	7,2	7,8	5,4

Tabelle 6 Ergebnisse der Metrik m_{Kt} für verschiedene Fragencorpora

3 Die Anwendung der Metrik

3.1 Überarbeitung des Fragenkatalogs

Den aktuellen Stand der Empfehlungen für Entwurfsrichtlinien haben wir in [SK16] zusammengefasst. Durch den Vergleich der strukturellen Charakteristiken von Fachfragen mit denen von Unterhaltungsfragen zeigt sich, dass diese Richtlinien im Wesentlichen berücksichtigt werden. Hauptsächliche Ansatzpunkte für eine Optimierung sind eine Reduktion der langen Wörter sowie der Wortanzahl in den Antworten [HR13]. Dabei ist jedoch der Lernkontext zu beachten: essentielle Informationen sollten nicht eliminiert werden.

Um zu prüfen, ob eine Modifikation der Fragen unter dieser Maßgabe zu einer geringeren textuellen Komplexität und damit verbunden, zu einer besseren Nutzungsmöglichkeit im Rahmen einer Quiz-App führen kann, wurde der Fragencorpus zum Thema *Anaerobtechnik* diesbezüglich bearbeitet. Wie in Tabelle 6 zu sehen ist, konnte durch die Änderungen der Wert für die Metrik mit 5,4 deutlich – fast auf das Niveau der Unterhaltungsfragen - verbessert werden.

3.2 Nutzerstudie

Mit Hilfe einer Nutzerstudie wurde die Wirkung der Änderungen überprüft: 14 Studierende des Studiengangs Umweltingenieurwesens spielten in der Lehrveranstaltung *Anaerobtechnik* hintereinander die drei QuizUp-Themen *Anaerobier Stout* (originaler Fragenkatalog), *Anaerobier Lager* (überarbeiteter Fragenkatalog) und *Lukas' Allgemeinwissen* für jeweils 20 Minuten gegeneinander. Es wurden zwei Gruppen gebildet, die die beiden Anaerobtechnik-Fragenkataloge in jeweils umgekehrter Reihenfolge spielten, um die Auswirkungen möglicherweise durch die Reihenfolge auftretende Effekte auszuschließen. Den Probanden wurden die Unterschiede zwischen den beiden Anaerobtechnik-Fragenkatalogen nicht erläutert. Nach jeder Spielphase wurden Fragebögen ausgegeben, in denen die Probanden zu ihren Eindrücken befragt wurden. Es sollten dabei die folgenden Hypothesen überprüft werden:

H1: Die Überarbeitung des Fragenkatalogs hat seine Verständlichkeit erhöht.

H2: Der überarbeitete Fragenkatalog senkt den Zeitdruck.

H3: Die Überarbeitung des Fragenkatalogs führt zu einem höheren Spielspaß.

3.3 Ergebnisse

H1. Die **Verständlichkeit** des Fragenkatalogs wurde mit Hilfe der Fragen, ob die Auswahlfrage präzise gestellt und inhaltlich korrekt sei, bewertet. Dazu wurden jeweils 5-Punkt Likert-Skalen genutzt, wobei fünf Punkte hohe Zustimmung signalisierten. Zwischen den beiden Anaerobtechnik-Fragekatalogen gab es nur geringe Unterschiede (3,32 vs. 3,39). Hingegen war der Wert für die Unterhaltungsfragen mit 4,19 wesentlich höher. Mit Hilfe der Überarbeitung konnte die Verständlichkeit daher nicht erhöht werden.

H2. Der auftretende **Zeitdruck** wurde mit den drei Items „Es war schwierig, die Fragen in der zur Verfügung stehenden Zeit zu beantworten“, „Ich hatte Probleme, die Fragen in der zur Verfügung stehenden Zeit inhaltlich zu verstehen.“ und „Der Zeitdruck hat dafür gesorgt, dass ich nicht sorgfältig gelesen habe.“. Auch hier wiesen die Unterhaltungsfragen mit einem Durchschnittswert von 2,56 den besten Wert auf. Die überarbeiteten Fragen waren zwar mit 3,19 deutlich schlechter, verursachten jedoch auch merklich weniger Zeitdruck als die Originalfragen (3,64). Die Überarbeitung sorgte daher für weniger Zeitdruck.

H3. Der erreichte **Spielspaß** wurde mit Hilfe der In-Game-Variante des *Game Experience Questionnaire* (GEQ) [IKP13] überprüft. Die Ergebnisse sind nicht eindeutig (s. Abbildung 2). Zwar zeigen sich für positiv mit Spielspaß verbundene Merkmale wie *Competence*, *Challenge* und *Positive Affect* gleichmäßig gute Werte, jedoch sind diese Werte nicht einheitlich besser für einen der Fragenkataloge. Die Werte für *Negative Affect* und *Tension* hingegen sind für die beiden Fachfragenkataloge erkennbar

schlechter als für das Allgemeinwissen. Demgegenüber fühlen sich die Studierenden in den Fachfragen kompetenter als beim Allgemeinwissen. Insgesamt ergeben sich für alle drei Themen spieltypische Ausprägungen der Merkmale, jedoch liefern diese Werte keine Grundlage dafür, Hypothese 3 anzunehmen.

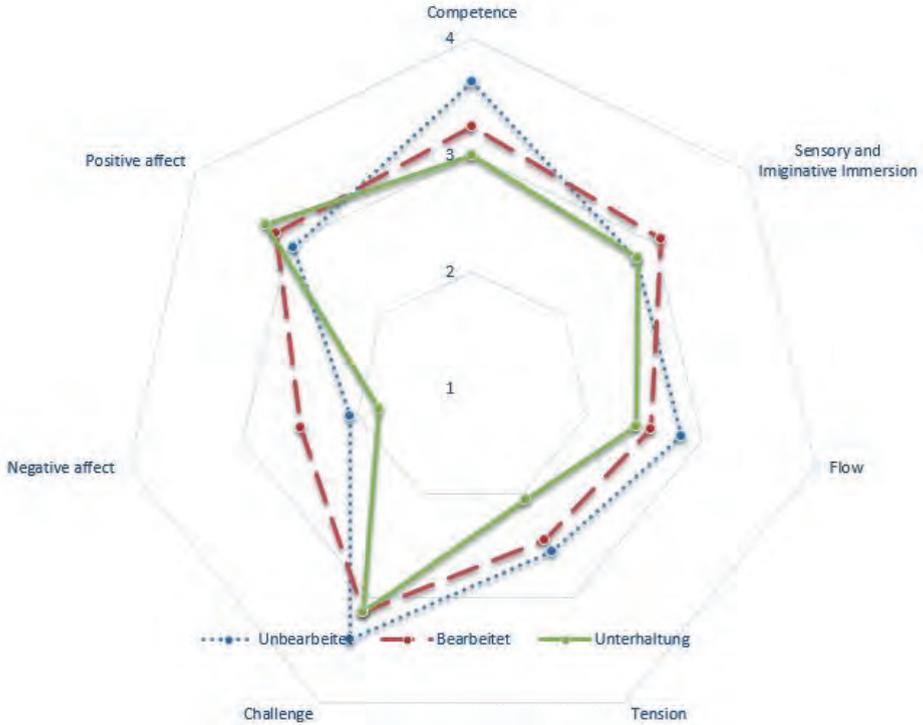


Abbildung 2 Ergebnisse GEQ

4 Diskussion und Schlussfolgerungen

Ziel der Arbeit war es, den offensichtlichen Popularitäts-Unterschied zwischen Fach- und Unterhaltungsfragen im Rahmen von Quiz-Apps aufzuarbeiten und insbesondere durch eine Analyse der textuellen Struktur zu einer höheren Attraktivität von Fachfragen beizutragen. Dazu wurde eine Metrik zur Messung der Komplexität einer Auswahlfrage entworfen, die weitestgehend automatisiert auf die Fragen angewendet werden könnte. Beim Entwurf der Metrik wurden u.a. Lesbarkeitsformeln genutzt um Hinweise auf die Optimierung der Fragen zu geben. Mit der Metrik konnten zum einen die strukturellen Unterschiede zwischen den beiden Fragenkategorien dokumentiert werden. Zum anderen wurden daraus Entwurfsgrundsätze erweitert, die insbesondere das Augenmerk auf

Gesamtlänge sowie Anzahl der Fach- und Fremdwörter legen. Ein diesbezüglich modifizierter Fragenkatalog erzielte beim Spiel in einer Quiz-App bessere Werte für den empfundenen Zeitdruck als der ursprüngliche Katalog. Die Werte eines Unterhaltungsthemas konnten jedoch nicht erreicht werden.

Dieses Ergebnis unterliegt sicherlich einer Reihe von Einschränkungen und Unsicherheiten. Die Stichprobe ist sehr klein, sodass bisher ausschließlich deskriptive Auswertungen möglich waren. Außerdem reichten die durchgeführten Fragenänderungen nicht aus, um bessere Ergebnisse in Bezug auf Spielspaß und Verständlichkeit zu erzielen. Während der Spielspaß mit der experimentellen Atmosphäre begründet werden könnte, deutet die Verständlichkeit eher auf grundsätzliche Spannungsfelder beim Fragenentwurf hin: durch eine versuchte Reduktion der Länge sowie der Fach- und Fremdwörter könnte die notwendige fachliche Präzision möglicherweise eingeschränkt werden. Inwieweit es möglich ist, in kurzen und prägnanten Worten einen für das Lernziel relevanten Sachverhalt auszudrücken, muss noch weiter evaluiert werden. So ist es zwar gelungen, die Fachfragen fast auf das Verständlichkeits-Niveau der Unterhaltungsfragen anzugleichen. Jedoch ist aufgrund des komplexeren Fachwissens nicht mit derselben Lesbarkeit zu rechnen. Gleichfalls ist zu prüfen, inwieweit durch die Vereinfachung auch ein Teil des Wissens verloren geht. Möglicherweise ist es notwendig, die Fragen in mehrere Fragen zu unterteilen, um dieselben Lernziele erreichen zu können, dabei aber durch einen schnelleren Lesefluss in der Quiz-Applikation einen höheren Spielempfinden zu erzeugen. Es ist einschränkend anzumerken, dass textuelle Komplexität nur eine von mehreren Charakteristiken ist, die für die Attraktivität der Fragen im Quiz-App-Kontext von Bedeutung sind. Zu den weiteren Faktoren zählen u.a. die Spielbereitschaft der Lernenden sowie die Inhalte der Frage selbst. Die Metrik als solche hat sich zwar als funktionierend erwiesen. Jedoch kann zwischen den Indikatoren der Anzahl von Fach- und Fremdwörtern sowie der Anzahl der langen Wörter ein Zusammenhang vermutet werden. Dies ist als Indiz zu werten, dass eine Überarbeitung der Metrik ihre Vereinfachung nach sich ziehen könnte. In einer folgenden Studie sollten der Einfluss textueller Komplexität bei Fach- und Unterhaltungsthemen auf den Lernerfolg sowie die Akzeptanz beider Themenbereiche untersucht werden.

Die dargestellte Metrik in Verbindung mit der Validierung hat jedoch einen Arbeitsstand erreicht, mit dem eine analytische Qualitätssicherung bei der Fragerstellung stattfinden kann. Durch eine Implementierung in einem Frageditor könnten Hinweise zur Steigerung der Lesbarkeit direkt bei der Fragerstellung gegeben werden.

Literaturverzeichnis

- [Am78] Amstad, T.: Wie verständlich sind unsere Zeitungen? Studenten-Schreib-Service, 1978.
- [BR07] Butler, A. C.; Roediger, H. L.: Testing improves long-term retention in a

- simulated classroom setting. *European Journal of Cognitive Psychology* 19, S. 514–527, 2007.
- [BV84] Bamberger, R.; Vanecek, E.: Lesen, verstehen, lernen, schreiben: die Schwierigkeitsstufen von Texten in deutscher Sprache. Jugend und Volk, 1984.
- [F148] Flesch, R.: A new readability yardstick. *Journal of Applied Psychology* 32, S. 221, 1948.
- [Gu52] Gunning, R.: *The Technique of Clear Writing*. McGraw-Hill, New York, 1952.
- [Ha14] Hardinghaus, B.: Interview mit LeBernd. Bernd Schneider über Quizduell, <http://www.spiegel.de/netzwelt/games/interview-mit-lebernd-bernd-schneider-ueber-quizduell-a-951529.html>, Stand: 11.02.2015.
- [Ha77] Halstead, M. H.: *Elements of Software Science (Operating and Programming Systems Series)*. Elsevier Science Inc, New York, NY, USA, 1977.
- [HR13] Haladyna, T. M.; Rodriguez, M. C.: *Developing and Validating Test Items*. Routledge, New York, 2013.
- [IKP13] IJsselsteijn, W. A.; Kort, Y. A. W. de; Poels, K.: The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. Manuscript in Preparation, 2013.
- [Mc76] McCabe, T. J.: *A Complexity Measure*, 1976.
- [P114] Plain Vanilla: QuizUp - Connecting people through shared interests, <https://www.quizup.com/>, Stand: 12.01.2016.
- [Po93] Poundstone, W.: *Prisoner's Dilemma/John von Neumann, Game Theory and the Puzzle of the Bomb*. Anchor, 1993.
- [Ru14] Russolillo, S.: QuizUp: The Next 'It' Game App?, <http://www.palmbeachpost.com/videos/news/is-quizup-the-next-it-game-app/vCYDgf/> <http://live.wsj.com/video/quizup-the-next-it-game-app/>, Stand: 28.01.2016.
- [SK16] Söbke, H.; Kämmerer, F.: Vermessene Fragen: Metriken als Ansatz automatisierter analytischer und konstruktiver Qualitätssicherung von Mehrfachauswahlfragen für mobile digitale Medien. In (Pfau, W. et al. Hrsg.): *Teaching Trends 2016-Vielfalt in der Lehre*, Clausthal-Zellerfeld, 10./11. November 2016. Waxmann, Münster, S. 153–162, 2016.
- [Sö15] Söbke, H.: Space for seriousness? Player Behavior and Motivation in Quiz Apps. In (Chorianopoulos, K.; al, E. Hrsg.): *Entertainment Computing – ICEC 2015 14th International Conference*, ICEC 2015 Trondheim,

Norway, September 29 – October 2, 2015 Proceedings. Springer, Cham, S. 482-489, 2015.

- [SW16] Söbke, H.; Weitze, L.: 2. The Challenge to Nurture Challenge Students' Perception of a Commercial Quiz App as a Learning Tool. In (Wallner, G. et al. Hrsg.): Entertainment Computing - ICEC 2016 - 15th International Conference, Vienna, Austria, September 28-30, 2016, Proceedings. Springer International Publishing, S. 15-23, 2016.

Evaluierung von E-Learning – Ein Kommentar zu „Media will Never Influence Learning“

Julian Dehne¹, Axel Wiepke² und Ulrike Lucke³

Abstract: Clark [C194] verfasste 1994 einen Artikel zu dem Thema „Medien haben keinen Einfluss auf das Lernen [eigene Übersetzung]“. Dieser Artikel wird als Grundlage genommen, um aus heutiger Sicht die Argumente von Clark zu thematisieren. Zunächst wird die Meinung von Clark zusammengefasst und mit aktuellen Publikationen aus der E-Learning Community in Deutschland verglichen. Daraufhin werden vier Paradigmen diskutiert, die sich aus den Argumenten ableiten lassen. Mit Hilfe dieser Paradigmen werden die DeLFI Publikationen der letzten Jahre hinsichtlich ihres Evaluationsfokus analysiert, visualisiert und interpretiert.

Keywords: Evaluation, Medientheorie, Paradigmen

1 Einleitung

Clark [C194] zeigte, dass bei der Betrachtung von Medien in Lernprozessen seit vielen Jahren die Methode mit dem Medium verwechselt würde. Er argumentierte, dass das Medium empirisch gesehen nie entscheidend für den Lernerfolg gewesen ist, sofern in der Evaluation auf eine alternative Methode hin kontrolliert wurde. Mit Methode ist dabei die pädagogische Methode (z.B. nach dem Berliner Model [HOS79]) gemeint. Er kontert das Argument, dass auch wenn Medien spezielle Funktionen bereitstellen würden (wie „Zoom“ für das Fernsehen), es immer auch andere Möglichkeiten gäbe, die gleichen Medienfunktionen mit herkömmlichen Mitteln abzudecken. Allerdings wird eingeräumt, dass Medien durchaus eine effizientere Art und Weise sein können, um eine pädagogische Methode anzuwenden. Wobei hier der Terminus „Effizienz“ die Kosten sowohl in Zeit als auch in Ressourcen dem Lernerfolg gegenüberstellt.

Dem entgegen steht die gängige Meinung (z.B. Schwill und Magenheim [MS12]), dass Lernerfolg das entscheidende Kriterium sei, um E-Learning-Forschung replizierbar und anfechtbar zu machen. Sie wenden sich in ihrem Beitrag bewusst gegen die Perspektive Schulmeisters, dass die Auswirkung vom Einsatz computergestützter Mittel auf den Lernerfolg kaum objektiv messbar sei, und schließen sich der empirisch orientierten pädagogischen Psychologie an. Im Sinne des psychologischen Forschungsparadigmas

¹ Institut für Informatik/Universität Potsdam, Lehrstuhl für Komplexe Multimediale Anwendungsarchitekturen, August Bebel Straße 19, 14482 Potsdam, julian.dehne@uni-potsdam.de

² Institut für Informatik/Universität Potsdam, Lehrstuhl für Komplexe Multimediale Anwendungsarchitekturen, August Bebel Straße 19, 14482 Potsdam, axel.wiepke@uni-potsdam.de

³ Institut für Informatik/Universität Potsdam, Lehrstuhl für Komplexe Multimediale Anwendungsarchitekturen, August Bebel Straße 19, 14482 Potsdam, ulrike.lucke@uni-potsdam.de

sei der digitale Einfluss auf den Lernerfolg messbar, es müssten jedoch die entsprechenden Variablen kontrolliert und Laborbedingungen bereitgestellt werden. Diese Diskussion halten wir für eine Betrachtung der etablierten Praktiken der E-Learning-Evaluation für aktuell.

2 Definition von Medien

Clark hält seinen Kritikern vor, dass diese als Lösung für ihre medienpädagogischen Ambitionen den Medienbegriff in die Methode hineindefiniert hätten. Dabei würde der Lernerfolg als Konsequenz der medienpädagogischen Neuerungen wie eine selbst-erfüllende Prophezeiung für die empirischen Studien wirken. Mit dem kritischen Rationalismus [Ke98] argumentiert ließe sich das Argument in die Formel übersetzen: „Kann mit meinem Evaluationsinstrument die Nicht-Wirkung, die Unerheblichkeit des neuartigen Mediums, belegt werden?“.

Um diesem Argument nachzugehen muss zunächst der Medienbegriff geklärt werden, um die Argumente von Clark übertragen zu können. Hier ließe sich mit der Digitalisierung und der Omnipräsenz von Computern argumentieren, dass die damals betrachteten Medien eine neue Qualität an Intelligenz gewonnen hätten und daher anders eingeschätzt werden müssen. Schiefner-Rohs & Hofhues [SH17] plädieren für einen weiten Medienbegriff, bei dem mediatisiertes Lernen als pädagogischer Raum zu verstehen ist, nicht als Werkzeug. Sie wenden sich dabei gegen das engere Konzept eines Werkzeuges, da diese austauschbar wären und nur durch einen Neuigkeits-Hype Bedeutung für das Lernen hätten, die eigentliche Struktur der Lernorganisation, die den erfolgreichen Einsatz desselben bedinge, jedoch meist ausgeblendet würde.

Die Argumentation von Clark bezieht sich hingegen auf einen engen Medienbegriff, bei dem ein spezifisches Werkzeug (anstatt einer vernetzten Medienaggregation) betrachtet wird. Dieser engere Medienbegriff ist für die E-Learning-Perspektive fruchtbarer, da hier einzelne Werkzeuge entwickelt werden, deren Wirkung von Interesse ist, das pädagogische Gesamtbild einer Institution jedoch eher vernachlässigt werden muss. Durch die Vernetzung muss genau genommen von medialen Ökosystemen gesprochen werden, die verbunden sind. Dadurch entsteht das Problem, dass die analytische Grenze eines Mediensystems zur Evaluation einzelner Elemente schwierig zu definieren ist. Die Grenze zwischen Medienraum und einzelner Werkzeug ist fließend und für jeden Fall muss einzeln entschieden werden, ob der Betrachtungsgegenstand ein Werkzeug für einen speziellen Fall darstellt oder einen Raum aufspannt.

Eine andere Frage besteht darin, ob der Medienbegriff eine andere Qualität von Werkzeug beschreibt, wenn wir von der heutigen Computertechnologie ausgehen. Aus der informatischen Perspektive dienen Medien zur Informationsverarbeitung, der Manipulation, Diffusion, Selektion, Generierung und Replikation von Informationen. Mit dieser Sichtweise stellen die heutigen Computer-basierten Werkzeuge keine neuartige Form von Medien dar, sondern unterscheiden sich von den klassischen Medien

(Film, Audio etc.) vor allem in der Komplexität, die durch die Computer-Eigenschaften entsteht. Daher gehen wir von einer pragmatischen Definition „digitales Artefakt zu dem Zweck der Verbesserung eines Lehr-Lern-Szenarios“ aus. Wenn wir im Folgenden zur besseren Flüssigkeit von „Tool“ oder „Medium“ reden, sei dies als Abkürzung zu verstehen.

3 Computerunterstützte Arbeitserleichterung als Effizienz im Lernprozess

3.1 Lernerfolg als Ergebnis orientiertes Paradigma

In Folge der PISA-Studie hat das Ergebnis orientierte Paradigma in der Pädagogik einen großen Aufschwung erhalten, da die Messung⁴ von schlechteren Lernergebnissen im internationalen Vergleich mit Prestigeverlust und einer grundsätzlichen Hinterfragung der bisherigen Praxen einherging. Als Konsequenz kommt auch der Messbarkeit von Lernsoftware eine große Bedeutung zu, wenn diese im Sinne der allgemeinen pädagogischen Psychologie als wirksam gelten soll. Hier greift jedoch der Einwand von Clark, dass das Medium nur Teil der pädagogischen Methode, aber nicht selbst wirksam sein kann. Zu direkter Lernwirksamkeit führt das Medium also nur, wenn es die Methode gegenüber allen anderen möglichen Methoden aufwertet, in dem es sie effektiver macht. Eine andere Argumentation wäre, dass ein Medium für einen methodischen Ansatz so entscheidend ist, dass es pars pro toto für die gesamte Methode stehen kann und somit das Portfolio an didaktischen Möglichkeiten bereichert. Damit würde die Stellung des Mediums oder Werkzeugs aufgewertet, dennoch hielte das Argument, dass nicht gezeigt wäre, dass andere Methoden nicht auch zu den gleichen Lernergebnissen führen könnten. Umgekehrt kann der Lernerfolg nicht aus einer Evaluation weggedacht werden, sobald politische Entscheidungen zu einem breiten Einsatz neuer Technologien ins Spiel kommen. Die Betrachtungsebene bei dem Lernerfolg muss demnach eine makroskopische sein, wohingegen bei diesem Paradigma die mikroskopischen didaktischen Entscheidungen, die zu dem Einsatz eines Werkzeugs in einem speziellen Setting führen, vernachlässigt werden. Die Kriterien, die ein Werkzeug nach diesem Paradigma beurteilen würden, folgen denen, die in dem Fach Psychologie etabliert sind. Denn dieses weist eine lange Tradition auf, interne Veränderungen wie kognitive Lernleistungen valide zu messen.

3.2 Gesteigerte Effizienz als Prozess orientiertes Paradigma

Clark liefert hier selber eine Lösung, indem er in einem Nebensatz anmerkt, dass das Medium die Effizienz einer Methode erhöhen könne, wobei nicht genau ausgeführt wird, was unter Effizienz zu verstehen ist. Gehen wir von der umgangssprachlichen Bedeutung

⁴ Auf die Methodendebatte zur PISA Studie soll hier nicht eingegangen werden.

aus, führt der Einsatz von Medien zu einem verringerten Ressourcenverbrauch (ob in Sachmitteln oder investierter Zeit). Die Idee ist die, dass die in der Software investierte Intelligenz zu einer Optimierung im Lehr-Lernprozess führt. Hier lassen sich schnell Beispiele benennen:

- Verbesserung der Kommunikation: Ein Wikieintrag kann von Vielen gelesen und editiert werden, ohne dass hierfür viel Papier bedruckt und ausgetauscht werden müsste.
- Verbesserung der Darstellungsmöglichkeiten: Eine Stadt im alten Rom kann visualisiert werden, das Sternensystem kann mit VR-Brille durchflogen werden etc.

Diesen Beispielen ist gemeinsam, dass Medien dazu führen den Prozess des Lernens zu verbessern, es jedoch nicht bewiesen werden kann, dass die Existenz des Wikis oder des geteilten Ordners dafür bürgt, dass eine höhere Lernwirksamkeit erreicht wird. Andere Faktoren (Bildungsgrad der Eltern, Intelligenz, ...) sind möglicherweise statistisch relevanter. Den Beispielen ist noch eine andere Eigenschaft gemeinsam, und zwar dass sie gewisse Funktionen in dem pädagogischen Prozess erfüllen, die diesen positiv beeinflussen. Diese Funktionen können analysiert werden, wobei es verschiedene Medien geben kann, die die gleichen Funktionen erfüllen. Zum Beispiel könnte der geteilte Ordner auch ein Papierordner sein, der herumgereicht wird, so dass jeder Lernende diesen für einige Zeit behält und mit seinen Notizen bereichert. Das entscheidende Kriterium ist hier, wie *effizient* ein Medium die pädagogische Funktion erfüllt, die von ihm erwartet wird. Das Effizienzkriterium stammt eher aus der Betriebswirtschaftslehre oder der Informatik, da diese Fächer die Traditionen haben, Kosten/Nutzen-Abwägungen eines Produktes oder einer Methode durchzuführen.

4 Gesellschaftliche Normalität und Akzeptanz als Prüfstein für innovative Tools

4.1 Das innovativ-künstlerische Paradigma

Software enthält neben der funktionalen Komponente auch eine künstlerische, in der Ästhetik wie auch der Innovation, dem Schaffen von Neuem. E-Learning-Software, die sich hieran ausrichtet, lässt sich jedoch nicht mit der klassischen psychologischen Methodik quantitativ bewerten, da die Möglichkeit der Kontrollgruppe fehlt. Eine genuin neue Entwicklung kann nur mit ihrer erfinderischen Kreativität punkten. Denn es gibt zwei Fälle: Entweder erfüllt die Software Funktionen, die so noch nicht realisierbar gewesen sind, so dass diese nicht vergleichend beurteilt werden können, oder sie aggregiert eine Reihe von Funktionen auf eine neue Art und Weise. Letzteres ist wieder ein Fall von gesteigerter Effizienz (siehe 3.2) und hat keinen plausiblen Effekt auf die Lernwirksamkeit.

4.2 Gesellschaftliche Akzeptanz, Usability und Normalität als Paradigma

Bei dem Übertragen der Argumentation von Clark in das Computer-Zeitalter muss eine andere Entwicklung ebenfalls berücksichtigt werden und zwar die gestiegene Entwicklungsrate. Da Computerprogramme keine Materialien benötigen und sich vielfältig manipulieren, verteilen und vor allem modular kombinieren lassen, ist es einfach und kostensparend, schnell neue Werkzeuge auf den Markt zu bringen ohne diese systematisch zu evaluieren. Dabei wird der Erfolg eines Produktes an der Annahmequote bemessen (oder anderen diffusen Innovations-Merkmalen). Dabei kann die Technologie sich überholen, bevor eine Evaluation überhaupt möglich ist. „Motor für Veränderungen sind dabei wechselseitig sowohl neue technische Entwicklungen (Web 2.0, Semantic Web, mobile Endgeräte, ...), neue Erkenntnisse zu Effekten und Einflussvariablen beim Lernen und Lehren und Änderungen in der bildungstheoretischen Auffassung vom Lernen“ [Dr13].

Fiedler⁵ argumentiert in einem persönlichen Gespräch, dass das Problem von Evaluationen im E-Learning-Bereich auf die Technik-Zentrierung und die fehlende Reflexion gesellschaftlicher Faktoren zurückzuführen sei. So sei die Digitalisierung nur eine der Kräfte, die auf die Gesellschaft wirkten. Andere Kräfte wären zum Beispiel die gestiegene Mobilität, die erhöhte Lebenserwartung etc. Damit würden der Digitalisierung Wirkungen auf die Gesellschaft und so auch auf die Lernsysteme zugeschrieben, die nicht von der Digitalisierung ausgingen. Während Clark die Evaluation bezüglich auf die Micro-Ebene kritisiert, bei der einzelne pädagogische Methoden ignoriert oder fehlinterpretiert werden, beschreibt Fiedler ein Missverständnis auf der Macro-Ebene, bei dem es um die Interpretation der „attributes of media“ geht. Dort, wenn man der Argumentation von Clark folgt, kann eine korrekte Evaluation von Medien nur dann erfolgen, wenn die besonderen Funktionen evaluiert werden, die dieses Medium gegenüber alternativen Medienzugängen hat. Diese Funktionen müssten aber hinsichtlich ihrer gesellschaftlichen Korrelate untersucht werden, wenn man Fiedler zustimmt. Die *Normalität* hat hier mehrere Bedeutungen. Zum einen geht es um die Normen, die ein E-Learning-Werkzeug erfüllen muss (Barrierefreiheit, Bedienbarkeit...). Zum anderen geht es um die Verbreitung und die damit implizierte Akzeptanz. Ein Werkzeug, was nicht benutzt wird, bringt in dem Sinne dieses Paradigmas keinen Nutzen. Eines, was allgegenwärtig ist, erfüllt keinen E-Learning-spezifischen Zweck. Damit ergibt sich eine zweidimensionale Matrix für die Einordnung von Medien:

⁵ Mitarbeiter HUL, Universität Hamburg - <https://www.hul.uni-hamburg.de/ueberuns/team/sebastianfiedler.html> - überprüft am 02.06.2017

Tabelle 1: Zeigt die Wechselseitigen Wirkung von Digitalisierung und der Effizienz eines Mediums

	geringer Einfluss der Digitalisierung	hoher Einfluss der Digitalisierung
hohe Effizienz des Mediums	Messbare Einflüsse, aber falsche Interpretation der Kausalität	Effizientes Erreichen des Lernerfolgs, was sich auch in den Ergebnissen niederschlägt
geringe Effizienz des Mediums	Unbeeinflusster Lernerfolg trotz Einsatz digitaler Ressourcen	Erreichen des Lernerfolgs bei großem Ressourcenaufwand

5 Ein aggregiertes Modell

Damit wurden vier Spannungsverhältnisse bei der Evaluation von digitalen Medien beschrieben (siehe Abbildung 1):

- Zunächst der in dem Clark Artikel beschriebene Fokus auf das Medium. Wenn ein digitales Tool als Medium betrachtet wird, muss dieses Funktionen besitzen, die gegenüber anderen verfügbaren pädagogischen Optionen zu einem höheren Lernerfolg führen. Lernerfolg als Output definiert, umfasst dabei auch so etwas wie gesteigerte Motivation. Wenn das Tool im Fokus steht, dann muss die Effizienz des Werkzeugs im Vergleich zu anderen Optionen mit gleichen Funktionen evaluiert werden. Wenn dagegen der Output im Vordergrund steht, dann wird die Rolle des Werkzeugs als Medium hinterfragt und ein Medienwechsel sollte experimentell die Wirkung eines Tools belegen.
- Das zweite Spannungsverhältnis besteht in der Messbarkeit. Desto innovativer ein Tool ist, desto größer ist der Entwicklungsaufwand und desto schwieriger ist es, Ressourcen und Vergleichsmöglichkeiten zu finden. Daher ist der Output „Lernerfolg“ diametral zu dem Design-basierten „Produkt als Innovation“ zu sehen.
- Das dritte Spannungsverhältnis ist soziologischer Natur. Wenn Wirkungen von digitalen Neuerungen betrachtet werden, muss die Gesellschaft als Kontext mitbetrachtet werden. Dabei kann es zur Verwechslung kommen, dass die Wirkung der Technologie von den Entwicklungen in der Gesellschaft überschattet wird. Der Innovationsgrad bemisst sich an der gesellschaftlichen Wirkung.

- Das vierte Spannungsverhältnis bezieht sich auf die Akzeptanz: Wenn eine Technologie so verbreitet ist, dass sie als Normalität gilt, dann ist es nicht möglich eine „Effizienz“ des digitalen Produktes zu messen, da dieses eine Monopolstellung einnimmt. Alternative Zugänge verlieren die Bedeutung, da es sich um kein digitales Artefakt mehr handelt, sondern um einen Alltagsgegenstand ohne wissenschaftliche Bedeutung. Umgekehrt, wenn das Werkzeug fremd für die Gesellschaft ist, wird es nicht akzeptiert und kann trotz theoretisch belegter optimaler Passung für ein pädagogisches Problem nicht zu dessen Lösung verwendet werden, ohne dass gesellschaftliche Änderungen vorausgehen.

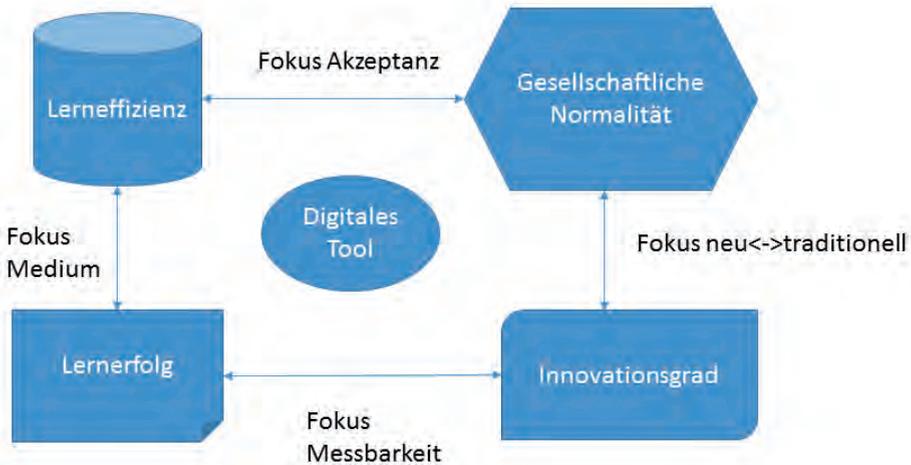


Abbildung 1: Vier verschiedene Paradigmen, die auf die Evaluation von digitalen Tools einwirken.

Dieses Modell kann dazu dienen, Evaluationen im Bereich E-Learning zu systematisieren und kann als Orientierung dienen, trotz der Vielfalt eine gemeinsame Beurteilungsgrundlage zu schaffen.

6 Inhaltsanalytische Anwendung des Modells

Um das Modell zu testen, wurden die DeLFI-Publikationen der letzten drei Jahre bezüglich der beschriebenen Paradigmen untersucht. Die Publikationen der ersten drei DeLFIs wurden zunächst ebenfalls analysiert. Da zu dieser Zeit die Evaluation noch kein fester Bestandteil der Forschungs- oder Publikationskultur war, wurde die Betrachtungsmenge auf die aktuelleren Publikationen eingengt. Als Methode wurde die quantitative Inhaltsanalyse gewählt. Folgende Variablen wurden gezählt:

- 1) Wurde eine Evaluation beschrieben? (ja-nein)
- 2) Wird eine einzelne Software/Medium/Tool diskutiert (ja-nein)
- 3) Geht es um ein konkretes Lehr-Lern-Setting? (ja-nein)

Weiterhin wurden die vier Paradigmen aus Abschnitt 5 wie folgt kodiert: Wenn ein Paradigma von den Autoren explizit verfolgt wurde (z.B. Messung des Lernerfolgs) wurde dieses mit 1 kodiert. Wenn das Paradigma implizit enthalten war oder einen geringen Stellenwert einnahm, wurde dieses mit 0 kodiert. Bei Nichterwähnung wurde eine -1 vergeben. Dabei wurde *die Qualität der unternommenen Evaluation ignoriert*. Es ging nur um die intendierten Evaluationsergebnisse.

- 4) Prozess orientiertes Paradigma (Effizienz): Wird die Effizienz betrachtet, die das Medium auf einen Prozess zu dessen Verbesserung ausübt? (-1,0,1)
- 5) Ergebnis orientiertes Paradigma (Lernerfolg): Wird der Lernerfolg betrachtet, den das Medium bewirkt? (-1,0,1)
- 6) Innovativ-künstlerische Paradigma: Wurde die Neuigkeit oder Ästhetik der verwendeten Technologien/Konzepte beleuchtet? (-1,0,1)
- 7) Akzeptanzorientiertes Paradigma: Wurde die Akzeptanz der Software in der Breite, die Usability oder die Normalität der Nutzung beleuchtet? (-1,0,1)

Daraufhin wurden die Ergebnisse danach gefiltert, dass die Variablen 1-3 mit ja beantwortet wurden. Die übrigbleibenden 28 Beiträge wurden auf einer in einem zweidimensionalen Koordinatensystem abgebildet, um die vertretenen Paradigmen zu visualisieren. Dabei stellt der Farbton die Gewichtung der Datenpunkte dar. Desto tiefer der Farbton, desto mehr Beiträge sind in diesem Bereich zu verorten. Dabei entspricht der Punkt (-1, 1) der Betonung des Mediums als Prozessverbesserung ohne Betonung der Verbreitung oder Akzeptanz desselben. Der Punkt (-1,-1) entspricht dem Ergebnis orientierten Paradigma. Der Punkt (1,-1) entspricht dem innovativen Paradigma und der Punkt (1,1) entspricht dem Paradigma der gesellschaftlichen Akzeptanz und Verbreitung. Dabei wurden die Variablen 4 und 5, 4 und 6, 5 und 7, und 5 und 6 je miteinander verrechnet. Damit entspricht der Punkt (0,0) einem Evaluationskapitel, bei

dem alle vier Paradigmen erwähnt wurden und in der Studie Berücksichtigung gefunden haben. Das Ergebnis ist in Abbildung 2 zu sehen. Weil die Annahme für die Visualisierung war, dass sich die Werte mit einem Polynom darstellen lassen, wurde die folgende Formel zur Darstellung der Verteilung mittels Gauß-Verfahren ermittelt:

$$f(x,y) = 0,175*(-x^2y + x*y^2 - x^2 + y^2 - 3x + 3y) \quad (1)$$

wobei x und y die sich gegenüberliegenden Paradigmen sind. Die Ergebnisse in Abbildung 2 zeigen, dass es drei Schwerpunkte bei den Paradigmen gibt.

- Es gibt einmal die innovativen Technologiebeiträge unten rechts.
- Es gibt die relativ ausgeglichen Lernergebnisbeiträge, bei denen implizit die Effizienz thematisiert wird.
- Es gibt die Praxisbeiträge, die näher an die Normalität rücken, ohne dabei auf empirischen Lernerfolg zu kontrollieren.
- Der Lernerfolg ist gegenüber der Verbreitung und den gesellschaftlichen Aspekten das dominierende Paradigma, jedoch lenkt je eines der anderen beiden Paradigmen (Innovation oder Effizienz) den Vektor (-1|1) zur Seite.



Abbildung 2: Zeigt die Verteilung der DeLFI Publikationen auf die Evaluationsparadigmen. Die dunkleren Kreise zeigen erhöhte Fallzahlen an.

7 Desiderate und Diskussion

Die Evaluation eines E-Learning Tools müsste je nach Schwerpunkt auf eine oder mehrere der folgenden vier Arten erfolgen:

- 1) Beurteilung, ob der Einsatz von digitalen Medien in diesem Bereich in Folge einer gesellschaftlichen Entwicklung geschieht, und damit die Validierung des Tooleinsatzes eine irrelevante Frage ist. Dies erfordert eine sozialwissenschaftliche Expertise.
- 2) Begründung, ob der Einsatz von digitalen Medien als Teil der pädagogischen Methode in dem vorhergesehenen Einsatzszenario einen Mehrwert bringt. Diese Begründung kann nicht empirisch erfolgen, da es hier um eine kreative Ausgestaltung von Unterricht geht, die von der Situation abhängt (Lerngruppe, Wissensgegenstand, Institutioneller Rahmen etc.). Aus Sicht eines Entwicklers müsste hier ein Einsatzszenario ausreichend beschrieben werden, um die Begründung des Einsatzes logisch zwingend zu machen. Dies erfordert eine pädagogische Expertise.
- 3) Beurteilung, ob die in 2 beschriebene(n) Funktion(en) von dem Tool effizient erfüllt wird. Das Tool sollte in dem beschriebenen Rahmen im Vergleich zu anderen Medien einen Einfluss zeigen. Mind. ein Medienwechsel sollte belegen können, ob die von dem Tool gewünschte Funktion auch wirklich erfüllt wird, oder ob durch den Einsatz des Tools künstliche Messartefakte entstehen Dies erfordert eine medienpädagogische Expertise.
- 4) Beurteilung, ob die innere Struktur des Tools für die gewünschten Funktionen optimal ist. Bei einer Neuentwicklung ist dies nur begrenzt möglich. Bei einer bestehenden Menge ähnlicher Tools (z.B. Clickersysteme) können diese bezüglich informatischer Kategorien vergleichend betrachtet werden (Usability, mathematische Korrektheit, Genauigkeit der Vorschläge etc.): Dies erfordert eine informatische/betriebswissenschaftliche Expertise.

Forschung im Bereich E-Learning ist inhärent interdisziplinär. „Sowohl die didaktische als auch die technische Konzeption von digitalen bzw. virtuellen Lernumgebungen bedingen unterschiedliche Perspektiven, u. a. die der Informatik, der Pädagogik, der jeweiligen Fachdidaktik(en) und der Psychologie.“ [Dr13]. Aus dieser Übersicht ergibt sich die Forderung nach interdisziplinären Teams in der Evaluation von E-Learning Werkzeugen, die eine tragfähigere Evaluation ermöglichen können. Sonst besteht die Gefahr, dass weiterhin Medien/E-Learning-Werkzeuge für Lernerfolge verantwortlich gemacht werden, obwohl ihr Einsatz entweder eine gesellschaftliche Tatsache (kein Forschungsergebnis) ist oder eine ineffektive Nutzung pädagogischer Ressourcen darstellt. Des Weiteren würde eine Positionierung der Autoren bezüglich der Paradigmen ihre Beiträge angreifbar und damit wertvoller machen, da die Kriterien für die Evaluation einer der etablierten wissenschaftlichen Traditionen folgten oder sie eine Evaluation mit guten Argumenten abgelehnten anstatt pro forma durchzuführen.

Literaturverzeichnis

- [Cl94] Clark, Richard E.: Media will never influence learning. *ETR&D* 42/2, S. 21–29, 1994.
- [Dr13] Drummer, J. et al.: Forschungsherausforderungen des E-Learning. In: (Breiter, Andreas; Rensing, Christoph, Hrsg.): *DeLFI 2013 - die 11. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)*. 8.-11. 9. 2013 in Bremen. Bonn: Gesellschaft für Informatik (GI-Edition Proceedings, 218), 2013.
- [HOS79] Heimann, P.; Otto, Gunter; Schulz, Wolfgang *Unterricht. Analyse und Planung*. 10., unveränd. Aufl. Hannover: Schroedel (Auswahl Reihe B, 1/2), 1979.
- [MS12] Magenheim, J.; Schwill, A.: Evaluation von E-Learning. *i-com* 11/1, S. 42–45, 2012.
- [Ke98] Keuth, Herbert (Hrsg.): *Karl Popper, Logik der Forschung*. Berlin: Akademie-Verl. (Klassiker auslegen, 12), 1998.
- [SH17] Schiefner-Rohs, M.; Hofhues, H.: Prägende Kräfte. Medien und Technologie(n) an Hochschulen. In (Othmer, J.; Weich, A.; Zickwolf, K., Hrsg.), *Medien, Bildung und Wissen in der Hochschule*. Springer Verlag, 2017 (eingereicht).

Digitalisierte Hochschuldidaktik: Qualitätssicherung von Prüfungen mit dem E-Assessment-Literacy-Tool EAs.LiT

Andreas Thor¹, Norbert Pengel² und Heinz-Werner Wollersheim³

Abstract: Die Formulierung von Learning Outcomes und deren Transparenz gegenüber Studierenden ist Grundlage für eigenverantwortliche Lernprozesse und kompetenzorientierte Prüfungen (Constructive Alignment). Das in diesem Beitrag präsentierte E-Assessment-Literacy-Tool (EAs.LiT) unterstützt hochschuldidaktisch fundiert bei der Formulierung von Learning Outcomes, der darauf basierenden Erstellung und Begutachtung von Aufgaben sowie der kriterienbasierten semi-automatischen Zusammenstellung gleichwertiger E-Prüfungen.

Keywords: E-Assessment-Literacy, Constructive Alignment, Qualitätssicherung, ILIAS

1 Einleitung

Die Verbreitung von E-Assessments an deutschen Hochschulen im formativen und summativen Bereich nimmt zu und ist ein wichtiger Aspekt zur Schaffung flexibler und transparenter Lernpfade [Ho15]. Dabei ist es notwendig, dass Lehrende das prüfungsdiaktische Potenzial von E-Assessments erschließen, um Lernprozesse und Leistungsüberprüfung kompetenzorientiert gestalten zu können (Constructive Alignment) [BT07]. Die Etablierung hochschulübergreifender Qualitätsstandards im Bereich E-Assessment benötigt dabei technische und personelle Unterstützungsstrukturen. So ist z. B. ein Austausch von Aufgaben (Items) zur qualitativen Begutachtung (Review) oder hochschulübergreifenden Verwendung in Prüfungen aufgrund verschiedener Learning-Management- und Prüfungssysteme bisher kaum möglich. Gleichzeitig ist aber der hochschulübergreifende Wissens- und Erfahrungsaustausch von Lehrenden entscheidend für die Entwicklung und Verbreitung von (*E-*)*Assessment-Literacy*. Darunter wird ein sowohl personell als auch organisatorisch-institutionell verankertes, strukturiertes Wissen u.a. um die Bedeutung und kompetenzorientierte Gestaltung von Prüfungen in Lehr-Lern-Arrangements verstanden, um die Qualität von Hochschulprüfungen zu sichern [WP16].

Dieser Beitrag präsentiert EAs.LiT, ein Web-basiertes System zur kollaborativen, qualitätsgesicherten Erstellung von Items für E-Assessments unter Berücksichtigung des Constructive Alignment, was im Rahmen eines vom Sächsischen Staatsministerium für Wissenschaft und Kunst geförderten Verbundprojekts entwickelt wurde. Dazu wurden die bisher in Learning-Management-Systemen (LMS) bekannten Elemente zur Item-Erstellung um die Definition von Learning Outcomes (LO) und Anforderungsstufen

¹ Hochschule für Telekommunikation Leipzig, thor@hft-leipzig.de

² Universität Leipzig, norbert.pengel@uni-leipzig.de

³ Universität Leipzig, wollersheim@uni-leipzig.de

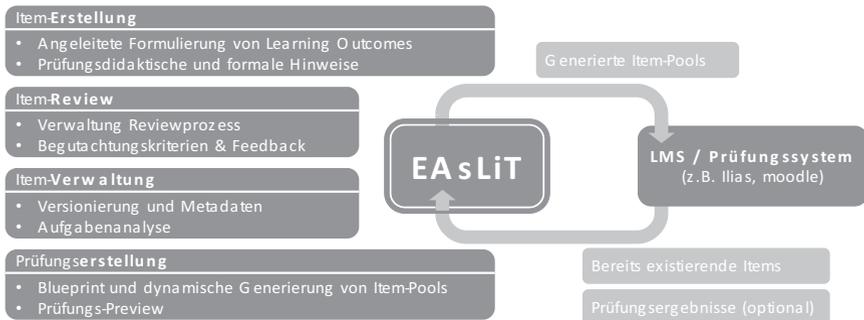


Abbildung 1: Schematische Darstellung des E-Assessment-Workflows mit EAs.LiT

sowie thematische Einordnung erweitert. Zusätzlich ermöglicht ein Begutachtungsprozess die Erstellung strukturierter Reviews zur Sicherung der Item-Qualität hinsichtlich fachlicher Richtigkeit, Relevanz und Formulierung. Des Weiteren unterstützt EAs.LiT die Erstellung von Item-Pools für Prüfungen durch eine interaktive, mehrdimensionale Datenexploration innerhalb des Item-Bestands basierend auf Strukturplänen für Prüfungen (Blueprints).

2 EAs.LiT-Workflow zur qualitätsgesicherten Item-Erstellung

Abbildung 1 illustriert den schematischen Workflow zur Item-Erstellung mit EAs.LiT. Dieser strukturiert den Prozess der Prüfungsentwicklung durch konkrete Arbeitsschritte und wird im Folgenden näher erläutert. Der gesamte Prozess wird während der Bearbeitung versioniert, sodass jederzeit Änderungen nachvollzogen und rückgängig gemacht werden können. Erstellte Items lassen sich als Item-Pool für verschiedene LMS exportieren und dort für Prüfungen verwenden; bereits in einem LMS vorhandene Items können importiert und in EAs.LiT weiterverarbeitet werden.

2.1 Erstellung von Learning Outcomes und Items

Die Erstellung von Items erfordert zunächst die Definition von Learning Outcomes (LO), d. h. was Lernende wissen, verstehen und in der Lage sind zu tun, nachdem sie einen Lernprozess abgeschlossen haben [Eu15]. Abbildung 2 (rechts) zeigt den Screenshot eines LO aus dem Bereich relationaler Datenbanksysteme zur Funktionalität von Löschregeln bei der Definition von Fremdschlüsseln. LO sind konkrete und messbare Kompetenzen, welche sich auf einen konkreten Inhalt beziehen, thematisch kategorisiert werden⁴ und jeweils einer Anforderungsstufe bzgl. der Lernzieltaxonomie nach Anderson und Krathwohl [AK01] zugeordnet sind. EAs.LiT unterstützt die Formulierung von

⁴ Die thematische Einordnung in eine Taxonomie ist aus Platzgründen nicht in Abbildung 2 dargestellt. Sie entspricht der Zuordnung von *Tags* oder Begriffen, wie man sie aus Webanwendungen kennt.

Fall- oder Problemvignette

Gegeben seien die Relationenschemata $R(a, b)$ und $S(c, d)$. Attribut b ist ein Fremdschlüssel auf S mit der Löschregel *ON DELETE CASCADE*.

Aufgabenstellung

Bei welchem der folgenden Ereignisse kommt die Löschregel zur Anwendung?

Antwort-Text	Punkte	Aktionen	
Löschen eines Tupels in R.	0	+	-
Löschen eines Tupels in S.	1	+	-
Einfügen eines Tupels in R.	0	+	-
Einfügen eines Tupels in S.	0	+	-

Learning Outcome

Löschregeln: Die Studierenden sind nach Abschluss der Lehrveranstaltung in der Lage, die Wirkung einer gegebenen Löschregel an einem konkreten Beispiel zu berechnen.

Löschregeln

Anforderungsstufe

	FW	KW	PW
1. Erinnern	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Verstehen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Anwenden	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
4. Analysieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Evaluieren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Erschaffen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abbildung 2: Beispiel eines Single-Choice-Items inkl. zugehörigem Learning Outcome (rechts)

LO durch Textbausteine, z. B. für Einleitungssätze („Die Studierenden sind nach Abschluss der Lehrveranstaltung in der Lage, ...“) sowie dem Bereitstellen von zur Anforderungsstufe passender Superverben (z. B. berechnen, prüfen oder übertragen für die Anforderungsstufe *Anwenden*). Zusätzlich zur Anforderungsstufe erlaubt EAs.LiT die Charakterisierung der Wissensdimension (Fakten-, Konzept- oder Prozesswissen), welche vom LO adressiert wird. Das in Abbildung 2 dargestellte LO wurde auf der Anforderungsstufe *Anwenden* im Bereich *Prozesswissen* eingeordnet.

Zur Erstellung eines Items wählt der Nutzer ein LO sowie einen Item-Typ aus. Der derzeitige EAs.LiT-Prototyp unterstützt *Single Choice* und *Multiple Choice*; weitere Item-Typen werden perspektivisch ergänzt. Abbildung 2 zeigt den Screenshot eines Single-Choice-Items (links) mit zugehörigem LO (rechts). Die Trennung des Fragebereichs in Fall-/Problemvignette und Aufgabenstellung erleichtert die Wiederverwendbarkeit in großen Item-Pools. So kann die Vignette aus Abbildung 2 auch für andere Aufgabenstellungen genutzt werden, um z. B. die Eigenschaften gültiger Attributwerte des Attributs b zu charakterisieren. Zusätzlich wird jedes Item einer Anforderungsstufe und Wissensdimension zugeordnet. Als Default-Wert wird beim Erstellen die Zuordnung des zugehörigen LO (rechts oben) verwendet. Abschließend wird jede Frage thematisch klassifiziert.

2.2 Begutachtung von Items (Review)

Der Item-Erstellungsprozess wird durch ein zweistufiges Peer-Review-Verfahren begleitet, um die Qualität der Items und damit des gesamten E-Assessments zu gewährleisten. Ähnlich dem Peer-Review-Verfahren bei wissenschaftlichen Zeitschriften geben die Lehrenden die von ihnen erstellten Items in den Begutachtungsprozess. Andere Nutzer (Peers) erstellen daraufhin ein strukturiertes, kriterienbasiertes Gutachten. Auf Basis der Reviews entscheidet ein Editor, ob das Item akzeptiert ist oder überarbeitet werden soll.

	Fachl. Richtigkeit	Relevanz bzgl. LO	Formulierung	Revisionsurteil
Fall- oder Problemvignette (alle gut)	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input type="radio"/> Item akzeptiert <input checked="" type="radio"/> Item überarbeiten <input type="radio"/> Item abgelehnt
Aufgabenstellung (alle gut)	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input type="radio"/> gut <input checked="" type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	
Antwortoptionen (alle gut)	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input checked="" type="radio"/> gut <input type="radio"/> Korrektur <input type="radio"/> ungeeignet	<input type="radio"/> gut <input checked="" type="radio"/> Korrektur <input type="radio"/> ungeeignet	
Feedback	<p>Die Frage nach dem Ergeignis prüft nur das Verstehen, nicht aber das Anwenden bzw. die Auswirkung. Der Begriff Löschregel ist ein Hinweis, dass die letzten beiden Optionen nicht korrekt sein können.</p>			Anforderungsstufe FW KW PW 1. Erinnern <input type="radio"/> <input type="radio"/> <input type="radio"/> 2. Verstehen <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> 3. Anwenden <input type="radio"/> <input type="radio"/> <input type="radio"/> 4. Analysieren <input type="radio"/> <input type="radio"/> <input type="radio"/> 5. Evaluieren <input type="radio"/> <input type="radio"/> <input type="radio"/> 6. Erschaffen <input type="radio"/> <input type="radio"/> <input type="radio"/>

Abbildung 3: Beispiel eines Item Reviews zum Item aus Abbildung 2

Abbildung 3 zeigt ein beispielhaftes Review für das Item aus Abbildung 2. Wesentlicher Bestandteil ist die Einschätzung der Item-Bestandteile Vignette, Aufgabe und Antwortoptionen bzgl. fachlicher Richtigkeit, Relevanz für das LO sowie Formulierung. Die Bewertung erfolgt dabei jeweils auf einer diskreten Skala (*gut* / *Korrektur* / *ungeeignet*). Analog wird für das gesamte Item ein Revisionsurteil (*akzeptiert* / *überarbeiten* / *abgelehnt*) vergeben. Zusätzlich fordert EAs.LiT ein konkretes Feedback zu den Kriterien, welche nicht als gut bewertet wurden. Im Beispiel von Abbildung 3 moniert der Gutachter u.a., dass die Anforderungsstufe niedriger ist als vom Item-Autor angegeben (*Verstehen* statt *Anwenden*). Diese Anpassung reflektiert der Gutachter durch seine Einordnung des Items in die Anforderungsstufe und Wissensdimension. Die vom Item-Autor vorgenommene Einordnung wird dabei farblich hinterlegt (siehe Abbildung 3 rechts unten), um Unterschiede schnell sichtbar zu machen.

2.3 Semi-automatische Generierung von Item-Pools

EAs.LiT stellt die Liste der erstellten Items inkl. wichtiger Attribute und Metadaten (u.a. Titel, Anforderungsstufe, Review-Status) übersichtlich in einer sortier-, filter- und durchsuchbaren Tabelle dar. Damit ist es Nutzern möglich, gezielt eine Menge von Items zu finden, ggf. zu bearbeiten und für einen Item-Pool zu verwenden. Darüber hinaus bietet EAs.LiT mit dem *Item Explorer* sowie dem *Blueprint Generator* Möglichkeiten, effektiv Item-Pools nach verschiedenen Kriterien zu generieren.

Der *Item Explorer* ermöglicht die interaktive Exploration einer großen Sammlung von Items. Dazu kann der Nutzer sich die Item-Anzahl nach verschiedenen Kriterien (Item Typ, Wissensdimension, Anforderungsstufe, thematische Kategorisierung sowie Schwierigkeitsgrad) in einer Pivot-Tabelle mittels Drag-and-Drop anzeigen lassen. Das Beispiel in **Error! Reference source not found.** zeigt, dass es derzeit zwei Items zum Thema *Normalformen* gibt, welche die Anforderungsstufe *Analysieren* haben. Ein Mausklick auf die Anzahl zeigt die dahinterstehenden Items in einer Voransicht. Die Exploration

kann auf beiden Achsen auf bis zu drei Ebenen verschachtelt werden. Würde der Nutzer nun z. B. noch zusätzlich den *Item Typ* auf die horizontale Achse neben *Anforderungsstufe* ziehen, würden die Werte noch nach *Single Choice* und *Multiple Choice* aufgegliedert.

Darüber hinaus bietet EAs.LiT die Möglichkeit der *Blueprint*-basierten Erstellung von Item-Pools. Hierzu definiert der Nutzer Bereichsbedingungen (minimale und maximale Anzahl von Items im Pool) bzgl. verschiedener Kriterien (u.a. Topic, Item-Typ, Anforderungsstufe), z. B. „zwischen zwei und fünf *Single-Choice*-Items“ oder „mindestens drei Items der Anforderungsstufe *Verstehen*“. Zusätzlich wird die Gesamtzahl der Items im Pool sowie das Ausmaß an Überlappungen zwischen Item-Pools definiert. EAs.LiT generiert daraufhin alle Item-Pools, welche allen Bedingungen genügen. Damit ist es möglich, (bedingungs-)äquivalente und hinreichend verschiedene Item-Pools automatisch generieren zu lassen, um insbesondere Prüfungen mit großen Kohorten in begrenzt-großen E-Assessment-Räumen durchführen zu können.

3 Implementation und Use Case

EAs.LiT ist als Plugin für das Content-Management-System (CMS) Wordpress implementiert. Dokumentation und Quellcode stehen auf GitHub⁵ zur Verfügung. Die wesentlichen Datenobjekte (LO, Items, Reviews) werden als spezielle Wordpress-Content-Typen in der von Wordpress verwalteten Datenbank abgelegt, welche somit auf standardisierte Wordpress-Funktionalitäten (Versionierung, Status- und Rechteverwaltung, thematische Kategorisierung) zurückgreifen können. Gleichzeitig ermöglicht die Verwendung von Wordpress-UI-Elementen eine adaptive, konsistente und moderne Benutzeroberfläche. Spezielle Sichten (u.a. Item Explorer), Algorithmen (u.a. Blueprint) sowie Import/Export-Funktionen zu LMS (derzeit QTI/Ilias 5) runden die Implementation ab.

Item Explorer

Item Typ Dimension Topic Level 2 Schwierigkeitsgrad

Anforderungsstufe

Topic Level 1

	Erinnern	Verstehen	Anwenden	Analysieren	Evaluieren	Erschaffen
Das Relationale Modell	1	3	1			
Normalformen		1	3	2	1	

Abbildung 4: Item Explorer zur Analyse der Items nach verschiedenen Kriterien

Der Prototyp wurde im Wintersemester 2016/17 für die Konzeption einer Modulprüfung für die Lehramtsstudiengänge an der Professur für Allgemeine Pädagogik der Universität Leipzig eingesetzt. Um knapp 600 Studierende unter Nutzung eines Computer-Pools mit 70 PC-Arbeitsplätzen zu prüfen, war es notwendig, 10 verschiedene, gleichwertige Prüfungen zusammenzustellen. Dazu wurden im Vorfeld die bereits vorhandenen Items vom Typ Single Choice und Multiple Choice aus ILIAS in EAs.LiT importiert. Darüber hin-

⁵ <http://www.easlit.de> bzw. <https://github.com/andreas-thor/eal>

aus wurden in EAs.LiT 20 neue Items erstellt und begutachtet. Aus den insgesamt 310 Items wurden durch einen Blueprint 10 Item-Pools generiert, die jeweils 20 Single Choice- und 5 Multiple Choice-Items enthielten. Gleichzeitig wurde sichergestellt, dass aus jeder Veranstaltungseinheit (zu gleichen Teilen Seminar und Vorlesung) mindestens ein Item stammt, wodurch in allen Prüfungsdurchgängen alle Themen geprüft wurden. Nach Sichtung dieser Items wurden sie als Fragenpools in ILIAS importiert und dort zusammen mit einer Freitext- und jeweils zwei Zuordnungs- und Anordnungsaufgaben zu 10 Prüfungen mit je 30 Items zusammengestellt. Die Nutzung von EAs.LiT ermöglichte nicht nur eine bessere Nachvollziehbarkeit der kriterienbasierten Zusammenstellung von Prüfungen, sondern auch (nach initialem Mehraufwand) eine deutliche Zeiterparnis. In den folgenden Durchgängen soll die Nutzung des *Blueprint Generators* um weitere Metadaten (u.a. LO und Schwierigkeitsgrad) erweitert werden.

4 Zusammenfassung und Ausblick

Ein wesentliches Potential digitalisierter Hochschulbildung liegt in der Nutzung von Technologien zur Vernetzung von Akteuren und Qualitätssicherung von Prozessen. EAs.LiT erlaubt eine vom Prüfungssystem unabhängige, hochschulübergreifende Kooperation in Aufgabenerstellung und –begutachtung. Es verbindet den hochschuldidaktischen Anspruch des Constructive Alignment mit der Usability eines modernen CMS. Um sowohl Erkenntnisse über die Bedeutung der *vor* der Prüfung festgelegten Metadaten (u.a. LO und Anforderungsstufe) als auch der im Rahmen einer Prüfung durch das Antwortverhalten von Studierenden generierten Kennwerte (u.a. Schwierigkeitsgrad, Trennschärfe, Reliabilität) für die Qualität einer Prüfung zu erlangen, sollen zukünftig geeignete Verfahren der Datenanalyse geprüft und in EAs.LiT implementiert werden.

Literaturverzeichnis

- [AK01] Anderson, L.W.; Krathwohl D. R. (Hrsg.): A Taxonomy for Learning, Teaching and Assessing. New York: Longman, 2001.
- [BT07] Biggs, J.; Tang, C.: Teaching for quality learning at university. What the student does. McGraw-Hill, Maidenhead, 2007.
- [Eu15] European Union: ECTS Users' Guide. Luxembourg: Publications Office of the European Union 2015, http://ec.europa.eu/education/library/publications/2015/ects-users-guide_en.pdf, Stand: 14.07.2017.
- [Ho15] Hochschulrektorenkonferenz zur Konferenz der europäischen Bildungsminister in Jerewan 2015, https://www.hrk.de/fileadmin/redaktion/hrk/news/_migrated/downloads/HRK_PM_Jerewan_15052015.pdf, Stand: 14.07.2017.
- [WP16] Wollersheim, H.-W.; Pengel, N.: Von der Kunst des Prüfens – Assessment Literacy. HDS. Journal – Perspektiven guter Lehre 2, S.14-32, <http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa-212985>, 2016.

Metriken zur Messung von Lernerfolg im Informatik-Grundlagen-Unterricht

Simon Rupp¹, René Pawlitzek¹ und Carlo Bach¹

Abstract: Es wird ein neu entwickeltes Werkzeug vorgestellt, das Metriken misst, mit denen der Lernerfolg der Studierenden in Übungsstunden abgeschätzt werden kann. Es handelt sich dabei um eine Erweiterung, die in der Entwicklungsumgebung BlueJ eingesetzt wird. Sie misst das Programmierverhalten der Studierenden anhand von Metriken, die aus Unterrichtsbeobachtungen und Dozierendenbefragungen hergeleitet wurden. Die Erweiterung berechnet aufgrund der gemessenen Metriken einen Hilfeindikator, der angibt, wie hoch das Bedürfnis nach Hilfe ist. Auf einer in Echtzeit generierten Webseite hat der Dozierende eine Übersicht über die gesamte Klasse. Alle Studierenden werden farblich, je nach Notwendigkeit einer Hilfestellung, eingefärbt. Vor Ort kann der Dozierende beim Studierenden über ein in die Entwicklungsumgebung integriertes Dashboard die Kennzahlen einsehen. Der Dozierende kann so seine Betreuungszeit auf sehr effiziente Art einsetzen und fördert damit die Wirkung des Unterrichts. Mit diesem Konzept werden potentiell die Studierenden unterstützt, welche die Hilfe am meisten benötigen. Die mit dieser Erweiterung gesammelten Daten wurden analysiert und führten zur Weiterentwicklung und Verbesserung der Metriken.

Keywords: Lernerfolg, Lernerfolgsmessung, Metriken, smart classroom, BlueJ

1 Einleitung

Im ersten Jahr des technischen Bachelorstudiengangs „Systemtechnik“ an der NTB, wird ein Grundlagenkurs in Ingenieur-Informatik im Umfang von 8 ECTS über zwei Semester geführt. Der Unterricht erfolgt im wöchentlichen Kontaktunterricht während der Vorlesungszeiten. Die Studierenden erlernen in einer Abfolge von kurzen Vorlesungsblöcken und betreuten Übungen ihre erste Programmiersprache. Während der individuellen Bearbeitung der Aufgaben am Rechner sind die Studierenden durchgehend durch einen Dozierenden betreut.

Das Ziel des Kurses ist, zukünftigen Maschinenbauern, Elektrotechnikern und Informatikern die Anwendung einer objekt-orientierten Programmiersprache zu vermitteln. Als Programmiersprache kommt Java zum Einsatz. Der Kurs verfolgt das Prinzip „Objects first“. Zum Einsatz kommen dabei das Buch „Java lernen mit BlueJ“ [BK09] sowie die Entwicklungsumgebung BlueJ [B117].

Während Unterrichtsbeobachtungen zur Analyse des Übungsteils hat sich gezeigt, was

¹ NTB Interstaatliche Hochschule für Technik Buchs, INF / PWO, Werdenbergstrasse 4, 9471 Buchs, Schweiz, simon.rupp@ntb.ch, rene.pawlitzek@ntb.ch, carlo.bach@ntb.ch

schon aus einigen Forschungsergebnissen bekannt ist: Der Dozierende kann sich bei Klassengrößen über 15 Studierenden nicht gleichzeitig auf alle Studierenden im Raum konzentrieren. Der Dozierende hilft häufig Studenten in seiner Nähe und übersieht dabei leicht andere.

Anders als ITS (intelligente Tutor-Systeme) [Kö07], unterstützende Analysen von Gruppenarbeiten [STL13] oder Grader wie der Praktomat [BHS16] wurde ein Werkzeug entwickelt, das den Dozierenden in den Übungslektionen dabei unterstützt, den hilfsbedürftigsten Studierenden zu identifizieren. Das Werkzeug misst dafür während des Bearbeitens der Übungen kontinuierlich das Programmierverhalten der Studierenden und bereitet es so auf, dass die Lehrperson die verfügbare Betreuungszeit effizienter einsetzen kann. Die Metriken werden autonom im Unterrichtsbetrieb gemessen und nicht anhand von Beobachtungen wie in [Hu06]. Einen ähnlichen Ansatz verfolgt Ulrich Kiesmüller bei der Untersuchung der Lösungsstrategien von Sekundarschülern. [Ki13]

2 Konzept

Bei jedem Studierenden läuft innerhalb der BlueJ-Entwicklungsumgebung eine Erweiterung, die Messdaten zum Programmierverhalten und zur Interaktion mit der Entwicklungsumgebung sammelt. Die Daten werden zu einem Indikator aggregiert und dem Dozierenden in Echtzeit auf einer Webseite zur Verfügung gestellt. Daraus kann der Dozierende ableiten, wo seine Hilfestellung den größten Mehrwert bringt. Die gemessenen Daten können in einem Dashboard abgerufen werden und geben Hinweise über die Aktivität und den Erfolg der einzelnen Studierenden.

Im Rahmen dieser Arbeit wurden Metriken entwickelt und verifiziert, die den größten Aussagegehalt für den Hilfeindikator bieten.

2.1 Metriken zur Messung von Lernerfolg

Um geeignete erfassbare Metriken zu finden, wurden Unterrichtsbeobachtungen und Dozierenden-Befragungen durchgeführt. An der Befragung nahmen fünf Dozierende teil. Aus den daraus gewonnen Erkenntnissen entstanden die zu messenden Metriken. Es handelt sich dabei um Aspekte, die von den Dozierenden entweder als wichtig beurteilt wurden oder von ihnen bei der Kontrolle von Quellcode angewendet werden. In Tabelle 1 ist ein Ausschnitt aus den zu messenden Metriken dargestellt. Die Metriken wurden anhand ihrer Wichtigkeit sortiert und entsprechende Maßnahmen eingeleitet.

Diese Arbeit beschreibt die grau hinterlegten Metriken (Tab. 1) und deren Ergebnisse. Die erste Version kann neben dem Messen der Interaktion der Studierenden mit der Entwicklungsumgebung auch Schlüsselwörter zählen und Kompilierereignisse interpretieren. Für die Quellcodeanalyse wurden mehrere statische Code Analyse Werkzeuge evaluiert und für die Verwendung in Version 2 vorbereitet.

Metrik	Wichtigkeit	Umsetzung in...
Quellcodeanalyse / Codier Konventionen	7	Vorarbeit / in Evaluation
Benutzung des Hilfefnopfs	5	Version 2
Fokus, Interaktion mit IDE	4	Version 1
Existenz von Schlüsselwörtern im Code	4	Version 1
Kompilermeldungen	2+	Version 1

Tab. 1: Metriken, die durch Dozierendenbefragungen und Unterrichtsbeobachtungen zusammengetragen wurden.

Kompilermeldungen: Gemessen wird die Anzahl Kompilervorgänge, die ein Studierender im Verlauf der Übungsphase durchführt. Diese Zahl kann in erfolgreiche und fehlgeschlagene Vorgänge aufgeschlüsselt werden. Bei Fehlgeschlagenen kann außerdem die Anzahl Fehler und Warnungen des Compilers ermittelt werden. Neben den Zählern, die sich von Beginn weg inkrementieren, gibt es abgeleitete Metriken wie „compiles per minute“. Diese setzt die Anzahl Kompilervorgänge ins Verhältnis zur Zeit, die für die Übung aufgewendet wurde. Anhand dieser Werte kann auf den Erfolg geschlossen werden: ist die Anzahl der fehlgeschlagenen Kompilervorgänge innert kurzer Zeit sehr hoch, versucht der Studierende vermutlich durch Ausprobieren ein Problem zu lösen. Dies wäre für den Dozierenden ein Indiz bei diesem Studierenden vorbei zu gehen und ihn bei der Problemsuche zu unterstützen.

Neben dem Zählen der Kompilervorgänge wird auch die Zeit dazwischen gemessen. Dafür wurden zwei Kennwerte entwickelt. Der Erste misst die Zeit zwischen zwei Kompilervorgängen, der Zweite misst die Zeit, bis ein erfolgreicher Kompilervorgang ausgeführt wurde. Neben der Zeitdauer kann aus dieser Metrik das Kompilerverhalten des Studierenden abgelesen werden: arbeitet der Studierende nach dem Versuchsprinzip oder kompiliert er erst nach langer Zeit ein großes, komplexes Programmstück?

Fokus, Interaktion mit IDE: Um die Interaktion mit der Entwicklungsumgebung zu messen, wird die Textzeigerposition in regelmäßigen Abständen ausgelesen. Dadurch kann in Erfahrung gebracht werden, wie oft sich diese innerhalb eines Texteditors oder zwischen mehreren Editoren verändert. Mit diesen Messdaten kann auf die Aktivität geschlossen werden.

Es wird zudem gezählt, wie viele Texteditoren offen sind und ob der Fokus auf der BlueJ Anwendung liegt. Über solche Messwerte kann eruiert werden, ob der Studierende aktiv die Übung bearbeitet. Sind viele Fenster offen und wechselt der Studierende sehr oft den Editor, kann dies ein Hinweis sein, dass der Studierende nicht weiss, wo er ein Stück Code umsetzen soll. Oder es kann ein Hinweis sein, dass er von einem anderen Programm Codestücke entlehnt.

Weitere Metriken messen die Anzahl Zeichen in einer Datei und die Schreibgeschwindigkeit der Studierenden im Editor. Damit soll gemessen werden, ob Studierende viel mit Copy-Paste arbeiten oder ob sie über längere Zeit nichts zu ihrem Quellcode hinzufügen.

3 Analyse

Die BlueJ-Erweiterung wurde in zwei Unterrichtseinheiten mit einer Klasse von 20 und 19 Studierenden getestet und ausgewertet. Die gesammelten Daten wurden periodisch in ein für den Dozierenden zugängliches Verzeichnis abgelegt. Diese Daten wurden für die anschließende Analyse verwendet. Während die Erweiterung in Betrieb war, wurde fortlaufend ein Indikator berechnet, welcher die Notwendigkeit einer Hilfestellung darstellte. Der Dozierende konnte während den Übungslektionen über einen Computer oder Tablet auf den Indikator zugreifen und erkannte unmittelbar, welcher Studierende seine Hilfe am dringendsten benötigte. Bei den Studierenden vor Ort konnte der Dozierende ein Dashboard öffnen, um sich einen Überblick über alle Metriken zu verschaffen.

3.1 Studierendenprofile

Die durch die Anwendung gesammelten Daten wurden verwendet, um Profile zu entdecken und daraus Schlüsse zum Lernerfolg zu erzielen. Die resultierenden Hypothesen gilt es in weiteren Testläufen zu überprüfen bzw. zu widerlegen. Die Profile wurden einerseits durch die grafische Visualisierung in Diagrammen und andererseits durch die tabellarische Darstellung der Messwerte entwickelt.

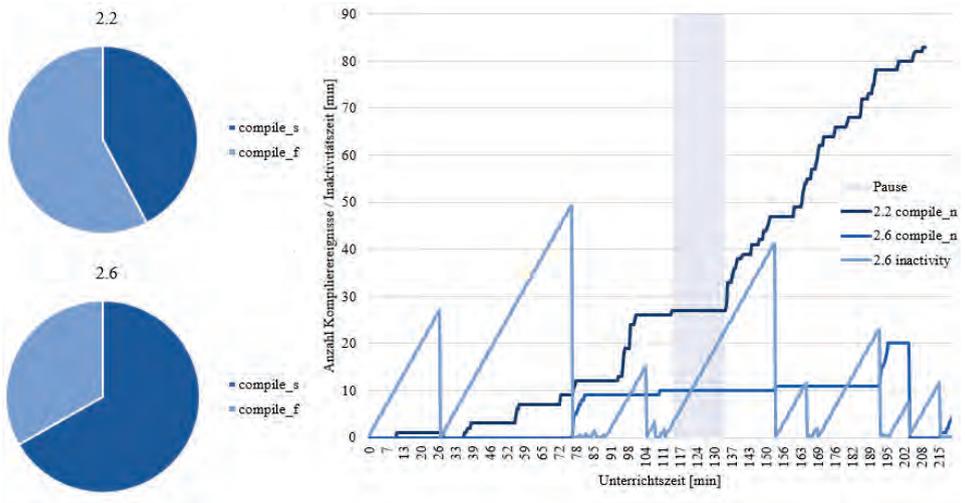


Abb. 1: Kreisdiagramm mit der Totalanzahl von erfolgreichen (compile_s) und fehlgeschlagenen (compile_f) Kompiliervorgängen (links) und Liniendiagramm mit der kombinierten Anzahl Kompiliervorgänge (compile_n) für zwei Studierende über den Zeitraum zweier Doppellektionen (rechts). Im Liniendiagramm ist außerdem die Inaktivitätskurve von Student 2.6 aufgetragen. Steigende Zacken zeigen die steigende Inaktivitätsdauer an. Grau hinterlegt ist die Pause.

In Abbildung 1 sind die beiden Studierenden 2.2² und 2.6 einander über den gesamten Verlauf der Übungsstunde gegenübergestellt. Der Studierende 2.6 hat mehrere Phasen von Inaktivität, was in dieser Grafik am waagrechten Verlauf der Kurve „compile_n“ und den hohen Peaks der Kurve „inactivity“ erkennbar ist. Durch das Betrachten der kleinen Anzahl Kompilervorgänge könnte darauf geschlossen werden, dass der Studierende 2.6 kaum programmiert und Motivationsprobleme hat. Im Diagramm links wird aber ersichtlich, dass der Studierende 2.6 meist erfolgreich kompiliert. Eine mögliche Schlussfolgerung wäre, dass er mit den Aufgaben unterfordert ist. Werden die erfolgreichen (compile_s) und fehlgeschlagenen (compile_f) Kompilervorgänge des Studierenden 2.2 in Relation zu seinem sehr häufigen Kompilieren gestellt, zeigt sich, dass dieser Studierende Unterstützung durch den Dozierenden benötigt.

Eine mögliche Klassifizierung der Studierenden in Gruppen ist in Tabelle 2 dargestellt. Die Messwerte wurden daraufhin untersucht, wie lange ein Studierender die Anwendung nutzte. Die Studierenden wurden in fünf Gruppen eingeteilt. Die größte Gruppe stellt die Gruppe von Studierenden dar, die zwei Doppellektionen durchgehend gearbeitet haben.

<50min	51-100min	101-150min	151-200min		>201min
1.3	1.14	1.16	1.1	1.11	2.2
1.6	1.19	2.11	1.5	1.12	2.6
1.15	2.13		1.7	1.17	2.9
	2.17		1.9	2.5	2.12
			1.10	2.14	

Tab. 2: Studierendengruppen anhand der Dauer in Minuten, wie lange die Studierenden die Entwicklungsumgebung nutzten.

<10min	11-20min	21-30min	31-40min	41-50min	>51min
1.3	1.14	1.5	1.1	1.9	2.7
1.6	1.19	1.12	1.10	2.6	2.16
1.15		1.17	1.11		2.9
		2.2	2.13		2.12
		2.5			2.14
		2.11			
		2.17			

Tab. 3: Studierendengruppen anhand der Dauer in Minuten, wie lange sie inaktiv waren.

Tabellenübergreifend wurden anschließend einzelne Studierende hervorgehoben, um weitere Schlussfolgerungen daraus zu erhalten. Werden Tabelle 2 und Tabelle 3 miteinander verglichen, kann erkannt werden, dass der Studierende 2.12 die Anwendung zwar sehr lange offen hatte, jedoch auch beinahe eine Stunde am Stück inaktiv war. Der Studierende 2.2 war hingegen nur solange inaktiv, wie die Pause lang war.

² Die Zahl vor dem Punkt gibt an, aus welcher Testgruppe der Studierende stammt. Die Zahl danach separiert die Studierenden innerhalb der Testgruppe.

4 Zusammenfassung und Ausblick

Als besonders aussagekräftig haben sich die Metriken rund um die Kompilier-Ereignisse und die Aktivitätsmessung gezeigt. Sieht der Dozierende den Bedarf für eine Hilfestellung, so kann er entsprechend reagieren.

In der nächsten Version sollen zudem einige Auswertungen nur die letzten Minuten repräsentieren und können pro Aufgabe zurückgesetzt werden. So wirken sich lokale Maxima oder Schwierigkeiten mit einer einzelnen Aufgabe nicht über den gesamten Übungsverlauf aus. Das Dashboard wird um eine Zeitkomponente ergänzt, damit auch vergangene Werte sichtbar gemacht werden.

Im Frühlingsemester 2017 wird die zweite Version der Erweiterung zum Einsatz kommen. Diese ermöglicht neben den bisher vorgestellten Metriken auch Quellcode zu analysieren und für die Auswertung zu verwenden. Testweise wurde die Bibliothek Checkstyle [Ch17] zur Überprüfung des Programmierstils eingefügt. Die Anwendung soll in größerem Umfang in mehreren Klassen getestet werden. Die während dieser Zeit gewonnenen Erkenntnisse können für die Weiterentwicklung und Verbesserung berücksichtigt werden. Das Ziel ist es, dieses Werkzeug flächendeckend im Informatik-Grundlagen-Unterricht einzusetzen.

Literaturverzeichnis

- [BK09] Barnes, D. J.; Kölling, M.: Java lernen mit BlueJ. Eine Einführung in die objektorientierte Programmierung. 4. aktualisierte Auflage, Pearson Studium (Informatik), München, 2009.
- [B117] BlueJ Java Development Environment, www.bluej.org, Stand: 20.02.2017.
- [Ch17] Checkstyle, www.checkstyle.sourceforge.net/, Stand: 20.02.2017.
- [Hu06] Hundhausen, C. D.; Brown, J. L.; Farley, S.; Skarpas, D.: A Methodology for Analyzing the Temporal Evolution of Novice Programs Based on Semantic Components. In ACM: ICER '06, Canterbury, United Kingdom, S. 59-71, 2006.
- [STL13] Slotta J. D.; Tissenbaum M.; Lui M.: Orchestrating of complex inquiry: Three roles for learning analytics in a smart classroom infrastructure. In ACM: LAK '13, Leuven, Belgium, S. 270-274, 2013.
- [Kö07] Kölle Ralph: Java lernen in virtuellen Teams. Kompensation defizitärer Rollen durch Simulation. 1. Auflage, Hülsbusch, Boizenburg, 2007.
- [BHS16] Breiter J.; Hecker M.; Snelling G.: Der Grader Praktomat, In (Bott O. J. et al., Hrsg.): Automatisierte Bewertung in der Programmierausbildung, 2016.
- [Ki13] Kiesmüller U.: Prozessbegleitende automatisierte Identifizierung von Problemlösestrategien Lernender beim Lösen algorithmischer Probleme mit Programmierumgebungen in der Sekundarstufe I. Shaker, 2013.

Beyond Lockdown: Towards Reliable e-Assessment

Bastian Küppers¹, Florian Kerber², Ulrike Meyer³, Ulrik Schroeder⁴

Abstract: In modern university education, lectures and accompanying exercises and tutorials incorporate digital elements to keep up with a general trend of digitalization. These digital elements spread from incorporating online learning management systems into the lectures to the usage of computers and tablets in exercises and tutorials. Despite e-Assessment being a valuable component in form of self-tests and formative assessment, the trend of digitalization has not yet been transferred on examinations. Retaining examinations on paper is often caused by reservations against e-Assessment, but also financial reasons interfere with the establishment of e-Assessment, because maintaining a suitable IT-infrastructure for e-Assessment is expensive in terms of money as well as administrative effort. Bring Your Own Device is a potential solution to this issue, but also poses new challenges regarding the integrity and reliability of examinations, hence potentially boosting the existing reservations. A common approach for securing e-Assessment is the usage of a so-called lockdown software on the students' devices, which is intended to ensure that students cannot execute impermissible actions during the examination. This paper exposes the disadvantages of current lockdown approaches in the Bring Your Own Device setting and outlines a novel alternative to securing e-Assessments. It thus contributes to reducing reservations against e-Assessment.

Keywords: E-Assessment, Digital Examinations, Bring Your Own Device, BYOD, Lockdown.

1 Introduction

Retaining examinations on paper is often caused by reservations against e-Assessment. These reservations concern for example the fairness or reliability of digital examination systems [VS09]. Especially the students have to accept e-Assessments before it can be introduced to higher education [FPH09, TE11]. Hence, when introducing e-Assessment to an institute of higher education, security is an often-discussed issue [VS09]. In any technical study program and in computer science programs in particular, the risk of fraud and cheating is presumably higher than for other study courses, because the particular students are studying the tool for e-Assessment - the computer. Thus, a proper mechanism to secure e-Assessment has to be used to ensure that it can be carried out reliably. Existing mechanisms, however, have some drawbacks especially in Bring Your Own Device (BYOD) settings. These include requiring administrative access rights to the BYOD

¹ RWTH Aachen University, IT Center, Seffenter Weg 23, 52074 Aachen, kueppers@itc.rwth-aachen.de

² RWTH Aachen University, IT Security Research Group, Mies-van-der-Rohe Str. 15, 52074 Aachen, kerber@itsec.rwth-aachen.de

³ RWTH Aachen University, IT Security Research Group, Mies-van-der-Rohe Str. 15, 52074 Aachen, meyer@itsec.rwth-aachen.de

⁴ RWTH Aachen University, Learning Technologies Research Group, Ahornstraße 55, 52074 Aachen, schroeder@cs.rwth-aachen.de

device and cross platform use on the usability side and a potentially incomplete threat model on the security side. In this paper, we therefore propose an alternative approach called LogDown. Our approach is based on a client-server architecture in which suspicious behaviour is logged and reported to the server. Thus, cheating during the eAssessment is discouraged as it will be detected. The rest of the paper is structured as follows: Related work is briefly discussed in Section 1.1., Section 2 details the drawbacks of lockdown software and Section 3 sketches our novel approach. We close with a conclusion and an outlook in Section 5.

1.1 Related Research

Several different approaches to ensure reliability already exist. For example online proctoring via remote-desktop software or surveillance over the webcam [Fr10]. These approaches induce, however, a lot of effort, because invigilators have to be available to review the remote desktop session or the webcam, keeping aside possible privacy issues. Therefore, a so-called lockdown software is most often used to prevent illegal actions during the e-Assessment [Fr10]. Once it has been started, this software controls the students' devices in a way that only certain actions can be carried out, e.g. starting a particular program or visiting white-listed webpages. Examples for existing software solutions are the Safe Exam Browser (SEB), which is an open source project developed at ETH Zürich⁵, and LockDown Browser, which is a closed source application by the company Respondus⁶. These tools, however, have some drawbacks, which introduce vulnerabilities that potentially render these tools attackable. While corresponding attacks require some technical effort and administrative access to the computer the tools are running on, e-Assessment on a centrally managed IT-infrastructure is not particularly vulnerable to these attacks. For BYOD e-Assessment, however, these vulnerabilities have to be overcome, since students usually have administrative access to their devices and time to prepare their devices before an examination.

2 Drawbacks of Lockdown Software

The drawbacks with lockdown software concern mainly two main aspects: usability and security. Both will be discussed in the following paragraphs. While the below mentioned drawbacks in usability could be acceptable, the drawbacks regarding security can potentially sabotage the fairness and reliability of e-Assessments.

2.1 Usability

Since lockdown software tries to control the working environment completely, it has to

⁵ <http://safeexambrowser.org>

⁶ <http://www.respondus.com/products/lockdown-browser>

interact with the operating system inevitably. Therefore, the software needs administrative privileges, at least while installing the software, in order to secure the working environment (see for example the manuals of SEB [ET17] and LockDown Browser [Re17]). In a BYOD setting, however, it cannot be guaranteed that students have administrative privileges on their devices. For example, a student employee, who is allowed to use a laptop provided by the employer, does not necessarily have administrative privileges on that device. Therefore, a lockdown application, which is intended to run on the students' devices flawlessly, must not require administrative privileges in order to establish a fully functional and secure BYOD scenario for e-Assessment.

Additionally, the available tools are not designed for cross-platform use. SEB supports Windows and MacOS, LockDown Browser can be used on iOS and ChromeOS additionally. To our knowledge, however, none of the available tools supports Linux, let alone all common platforms. Therefore, some students may not even be able to use a lockdown software on their accustomed operating system.

2.2 Security

Lockdown software attempts to provide a secure working environment by preventing potential security threats. Since SEB is open source, it is easy to get an insight into the methodology that it uses to implement this strategy. Therefore, SEB will be discussed here, but the results seem equally valid for other available lockdown software, because the feature set for that type of software is similar for all available products [FL13].

SEB uses many of the possibilities that MS Windows offers in order to secure the working environment. The implemented measures range from creating a new desktop object, which does not provide the default GUI, up to the use of windows hooks in order to get aware of occurring events. This approach is a negative security approach [Tr09], i.e. a threat model has been defined and for each threat, a counter measure was implemented to prevent the threat. The general problem with negative security is that it does not help to build trust, because it is hard to prove that every potential threat is covered in the threat model [As13].

Indeed, it seems that the threat model of SEB does not cover all potential threads [Sø16]. In fact, we carried out some tests ourselves and were able to inject a DLL file into SEB by using common DLL injection techniques [HB99]. The injected DLL allowed us to potentially execute any code. We used it to open a new window that displays information about the topic of the examination as a proof-of-concept. SEB could of course include our particular approach into the threat model, but that does not solve the problem in general.

3 An Alternative Approach: LogDown

Due to the previously discussed drawbacks of lockdown software using the example of SEB, we conclude that lockdown software does not offer suitable security for e-Assessment, especially not in a BYOD setting. Therefore, we propose an alternative

approach, which makes use of a client-server architecture. Our approach does not try to prevent cheating, but to detect suspicious behaviour, which is then reported to the server (hence the name 'LogDown'). Therefore, no deeper embedding into the operating system is necessary and thus no administrative privileges are required. Additionally, since there is only a loose coupling with the operating system, porting LogDown to a new platform is a lot easier.

LogDown is designed in a way that it only monitors itself, not the whole working environment. Broadly speaking, it checks if any communication channel, e.g. keyboard or screen, shows suspicious behaviour. To accomplish this goal, several things have to be checked, for example running processes or the contents of the clipboard. As opposed to using a threat model, this approach uses more of a safety model.

In order to make the approach taken with LogDown work, it has to be ensured that an unaltered version of LogDown is running on the students' devices, otherwise attacks like DLL injections or code injection in general would render LogDown useless. Since the students' devices are untrusted platforms from the examiner's point of view, there is virtually no possibility to ensure that students have not altered the LogDown executable when they are starting it. Since the executable could be altered, implementing a self-check, which is executed at the start, is pointless. It is, however, possible to verify that the executable has not been altered at run-time.

In order to do that, several steps have to be taken. First, every started instance of LogDown has to register itself at the server, which in turn provides the assignments of the examination. Hence, no registered version of LogDown means no assignments. In order to be able to match the registered instances of LogDown with the corresponding students, public key cryptography is used. All students have to register their key-pair prior to the examination. That key-pair is then used to establish a TLS connection with client authentication between LogDown and the server. This connection is held active throughout the whole examination and takes up the role of a heartbeat measurement. Should this connection break down for any reason, this is immediately reported to the examiner, who can then take action as needed and e.g. instantly stop the examination for the student in question. This is necessary, because a breakdown of the connection could have also different reasons than a cheating attempt, e.g. a problem with the Wi-Fi connection. Therefore, an automated termination of the examination could be wrongful.

To verify the integrity of the client, the server has a set of verification binaries. These verification binaries take up the role of a challenge-response authentication between the server and the LogDown executable. At random intervals, the server selects a random verification binary as challenge and sends it to the client to be executed with the expectation to receive a timely answer. A late or wrong answer as well as no answer at all will result in the immediate notification of the examiner, as this possibly indicates a tamper attempt. Restricting the time for an answer to the server and sending out the verification binaries in random time intervals is necessary, as the verification binaries provide integrity checking code as well as the answer of the challenge, which is sent back to the server in

case of a successful check. Without these measures, analysing the binary and extracting the correct answer without an actual genuine LogDown executable would become feasible. To prevent pre-computations of challenge results as well as raising the bar for potential attackers, the verification binaries could be personalized or compiled just-in-time with random optimizations or code fragments. This way, a potentially compromised client would be unable to collect all possible verification binaries. Additionally, a successful verification allows reusing the verification binary in question, as the integrated self-check ensures the sanity of the platform. Therefore, no verification binaries are collected. A failed test could indicate a potential security breach and therefore the leak of the binary used, but due to the massive amount of potential binaries as well as the knowledge of the actual actor, due to identification via the TLS certificates, the number of attempts is restricted to one per person. In order to prevent leakage of the contents of the examination, the assignments are only sent to the students' device after the first successful execution of a verification binary. Figure 1 shows the workflow of LogDown.

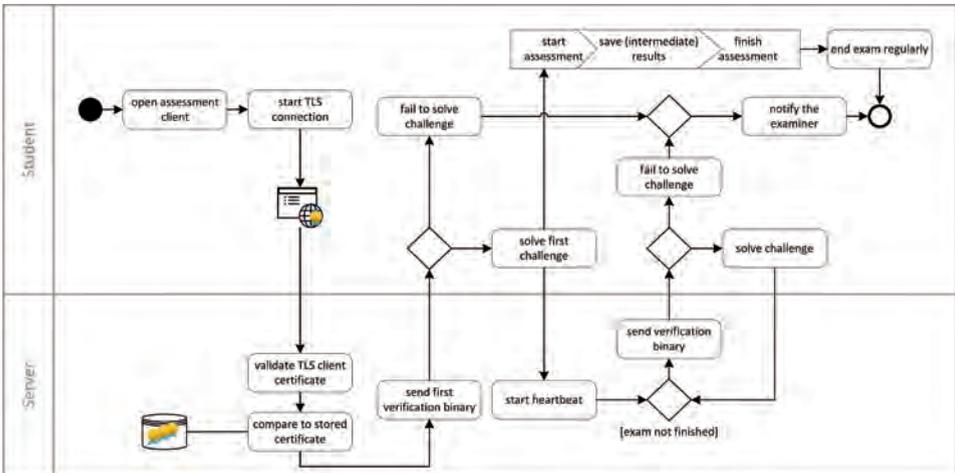


Fig. 1: Workflow of LogDown.

Unlike other hacking scenarios, where the aggressor can act out of anonymity, our approach removes this cover and therefore creates a psychological inhibition threshold, since even a single unsuccessful attempt to cheat is detected and can result in consequences for the student. Therefore, we expect that the willingness to cheat in an examination will be low in comparison to a regular lockdown application, despite the students' not being constantly monitored. Since the students can never know whether a verification binary is executed at a time or when the next execution is due, cheating is very risky. That situation resembles the state-of-the-art in paper examinations, where students could have crib sheets, but would risk to be caught by an invigilator when using these.

4 Summary and Future Work

In this paper, we discussed the general approach of conventional lockdown applications for e-Assessment systems, and pointed out their drawbacks and insecurities, especially in a BYOD scenario. Based on the results of this discussion, we proposed an alternative approach using a monitoring client-server architecture, which utilizes integrity checks at random time intervals. This does not only decrease the intrusiveness of the e-Assessment, it also increases the overall security due to its on-the-fly self-checks. Therefore, our approach potentially leads to a higher acceptance of e-Assessment and BYOD. Our next steps will be to implement the presented approach prototypical and run test in order to become aware of potential pitfalls. Especially the portability for different, common operating systems will be an issue that has to be addressed while implementing and testing.

Bibliography

- [As13] Association, Information Resources Management: Crisis Management: Concepts, Methodologies, Tools and Applications. IGI Global, Hershey, PA, USA, 1st edition, 2013.
- [ET17] ETH Zurich, Educational Development and Technology (LET): Windows User Manual. Website, 2017, http://safeexambrowser.org/windows/win_usermanual_en.html, Stand: 17.07.2017.
- [FL13] Foster, David; Layman, Harry: Online Proctoring Systems Compared, 2013.
- [FPH09] Fluck, Andrew; Pullen, Darren; Harper, Colleen: Case study of a computer based examination system. *Australasian Journal of Educational Technology*, 25/4, 2009.
- [Fr10] Frank, Ariel J.: Dependable distributed testing: Can the online proctor be reliably computerized? In (Marca, David A., Hrsg.): *Proceedings of the International Conference on E-Business*. SciTePress, S.1, 2010.
- [HB99] Hunt, Galen; Brubacher, Doug: Detours: Binary Interception of Win32 Functions. In: *Third USENIX Windows NT Symposium*. USENIX, S. 8, July 1999.
- [Re17] Respondus Technical Support: Knowledge Base. Website, 2017, <http://support.respondus.com/support/index.php?/Knowledgebase/Article/View/147/25/i-need-administrator-rights-to-install-respondus-lockdown-browser-windows>, Stand: 17.07.2017.
- [Sø16] Søgaard, Thea Marie: Mitigation of Cheating Threats in Digital BYOD exams, 2016.
- [TE11] Terzis, Vasileios; Economides, Anastasios A.: The acceptance and use of computer based assessment. *Computers & Education*, 56/4, S.1032–1044, 2011.
- [Tr09] Trost, R.: *Practical Intrusion Analysis: Prevention and Detection for the Twenty-First Century*. Pearson Education, 2009.
- [VS09] Vogt, Michael; Schneider, Stefan: *E-Klausuren an Hochschulen: Didaktik - Technik – Systeme - Recht - Praxis*, 2009.

Digitale Signierung elektronischer Prüfungen: Organisation und Arbeitsumgebung im Prüfungsraum

Elisabeth Katzlinger-Felhofer¹ und Martin Stabauer¹

Abstract: Elektronische Prüfungen gewinnen in den letzten Jahren an vielen Hochschulen und anderen Bildungseinrichtungen stark an Bedeutung. An allen Fronten wird an Lösungen gearbeitet, diese effizient und für alle Beteiligten schnell, einfach und sicher abzuhalten. An der Johannes Kepler Universität Linz wurde im Sommersemester 2011 ein elektronischer Prüfungsraum eingerichtet und seither ständig weiterentwickelt. Heute können auf äußerst benutzerfreundliche und hoch automatisierte Art Prüfungen aller Fachbereiche abgewickelt werden. Organisatorisch sicher verankert und unter Einbeziehung digitaler Signaturen auf den Studierendenausweisen zur rechtlichen Absicherung kann diese Lösung als einzigartig im deutschsprachigen Raum gesehen werden. Diese Forschungsarbeit diskutiert die bisherigen Entwicklungen, zeigt einen Überblick über den Status Quo und gibt einen Erfahrungsbericht über die vergangenen sechs Jahre.

Keywords: Elektr. Prüfung, Prüfungsraum, Digitale Signatur, Bürgerkarte, Moodle, E-Assessment

1 Einleitung

Der vorliegende Beitrag beschäftigt sich mit der organisatorischen, räumlichen und technischen Ausstattung eines elektronischen Prüfungsraumes an der Johannes-Kepler Universität (JKU) Linz, der den aktuellen didaktischen, studienrechtlichen und (sicherheits-) technischen Anforderungen entspricht. Es wird auf sechs Jahre Erfahrung mit elektronischen Prüfungen zurückgeblickt. Eine Befragung der betroffenen Lernenden zum Arbeiten im Prüfungsraum wurde durchgeführt und die Ergebnisse in diesem Beitrag dokumentiert. Auf didaktische Ansätze wird nicht näher eingegangen, sie sind aber Grundlage für die Planung und Gestaltung der Prüfungsprozesse.

Mit der Reformierung der Studienstrukturen im Zuge des Bologna-Prozesses steigt die Prüfungsbelastung der Lehrenden und der Wunsch nach standardisierten Prüfungen, die die Vergleichbarkeit von Bildungsabschlüssen im Hinblick auf die erreichten Kompetenzen erhöhen. So bieten Universitäten unterschiedliche Formen der von E-Assessments an [Ha16], [KB16], [BEH09]. E-Assessmentsysteme verwenden unterschiedliche Frage- und Aufgabenformen, die automatisch ausgewertet werden. Je nach Fachgebiet stehen angepasste Aufgabenformen zur Verfügung, wie bspw. der an der JKU entwickelte Fragetyp „Buchungssatz“ für Buchhaltung oder aber auch „open book exams“, bei dem die Studierenden Rechercheaufgaben in Form von Webquests lösen. Die (teil-) automatisierte Korrektur der Prüfung spielt dabei eine zentrale Rolle [Ko16].

¹ Johannes Kepler Universität, Institut für Digital Business, Altenberger Straße 69, 4040 Linz, Österreich, {elisabeth.katzlinger, martin.stabauer}@jku.at

Die Umstellung von Prüfungsprozessen auf computerbasierte Verfahren bringt neue Herausforderungen für die Prüfungsorganisation, -konzeption, -methodik, -technik und -systeme sowie dem Prüfungsrecht mit sich. Die erhoffte Ressourceneinsparung (von Zeit und Personal), die Aussichten auf eine Standardisierung sowie die einfache Auswertung von Tests sprechen für den Einsatz von E-Assessment Tools [Fi13].

2 Organisation

Die Einrichtung des Prüfungsraumes stand unter der Prämisse, den Studierenden mehr Prüfungstermine und kürzere Wiederholungszyklen anzubieten, um so einen Beitrag zur Studienbeschleunigung zu leisten, ohne die Prüfenden zu belasten. Die Prüfungsaufsicht übernimmt administratives Personal, für etwaige IT-technische Fragen ist ein Techniker erreichbar. Um unklare Situationen zu vermeiden, müssen detaillierte schriftliche Anweisungen und sorgfältig entwickelte Aufgabenstellungen vorhanden sein [Lo16].

Die JKU Card ist ein Lichtbildausweis für Studierende der JKU. Sie dient nicht nur der Identifizierung bei Prüfungen, sondern auch einer Vielzahl anderer Aufgaben wie etwa der Einfahrt auf den Parkplatz, bargeldloser Bezahlung am Campus und der Mensa, Entlehnung von Büchern in den Bibliotheken, Druckfunktionen und andere. Auf ebendieser Karte können Studierende kostenlos die Bürgerkartenfunktion (siehe folgendes Kapitel) aktivieren lassen. Um den Immatrikulationsprozess zu beschleunigen, wird diese Aktivierung erst später durchgeführt. Da die JKU offizielle Registrierungsstelle ist, kann diese durch speziell geschultes Personal im Bereich der Prüfungsräumlichkeiten vorgenommen werden, dies dauert etwa 10-15 Minuten pro Aktivierung.

In Österreich sind zahlreiche Funktionen des E-Government online verfügbar, wie etwa Steuerausgleich, Versicherungsdatenabfrage, Strafregisterauszug und ähnliche. Zu deren Nutzung bedarf es einer rechtsgültigen elektronischen Unterschrift, welche durch die sogenannte Bürgerkarte² geleistet wird. Diese ist in zwei Formen verfügbar: Zum einen kann sie auf Chipkarten, wie der e-Card (die Patientenkarte der österreichischen Sozialversicherungsträger) oder der JKU Card, aktiviert werden. Zum anderen die sogenannte Handy-Signatur via App oder SMS. An der JKU wurde die für Studierende kostenlose Variante der Karte gewählt, da alle Studierenden automatisch die passende Karte erhalten, während Mobiltelefone bei den Prüfungen nicht erlaubt sind. Die Bürgerkarte ermöglicht rechtsgültige Unterschriften durch die Signierung von PDF- und anderen Dokumenten. Das hierzu nötige Zertifikat kann durch die auf den PCs im Prüfungsraum installierte PKI-Umgebung (Public Key Infrastructure) überprüft werden. Zu diesem Zweck wurde ein Modul zur Integration der Bürgerkartenumgebung in Moodle entwickelt, welches selbst auf Dienste der Signaturerstellung nach dem österreichischen E-Government-Standard zurückgreift³.

² <https://www.buergerkarte.at>

³ <https://www.buergerkarte.at/downloads-karte.html>

3 Prüfungsraum und -ablauf

Im elektronischen Prüfungsraum sind 62 Prüfungsplätze aufgebaut, davon sind zwei rollstuhlgerecht ausgeführt (breitere und höhenverstellbare Tische). Jeder Prüfungsplatz besteht aus einem handelsüblichen PC, der vor Manipulation sicher in einem Fach des Tisches eingesperrt ist. Er kann ohne Schlüssel nicht von außen bedient werden, auch können keine USB-Sticks oder ähnliches angesteckt werden. Das An- und Ausschalten erfolgt zentral über einen gesonderten Steuerungs-PC.

An die PCs der Prüfungsplätze angeschlossen sind unterhalb der Tischplatte aufgehängte Bildschirme, die durch eine in diese eingelassene Glasplatte sichtbar sind. Durch diesen Aufbau nach [Hu10] können die Studierenden von nur wenigen Aufsichtspersonen überblickt werden. Außerdem ist damit gewährleistet, dass niemand auf die Bildschirme in der Reihe vor sich sehen kann. Zwischen den einzelnen Arbeitsplätzen ist zwar eine 20 cm hohe Trennwand angebracht, die aber nur das Verschieben von Materialien auf den benachbarten Arbeitsplatz verhindern sollen. Mit etwas Verrenkung könnte man dennoch auf den Bildschirm des Nachbarn sehen, daher ist auf den Bildschirmen eine polarisierende Folie angebracht. Auf der Tischplatte ist Platz für Maus, Tastatur mit Chipkartenleser und Unterlagen, sofern bei der jeweiligen Prüfung erlaubt sind. Die PCs werden mit einer aktuellen Version von openSUSE Linux betrieben.

Für die Studierenden ist ein Wartebereich außerhalb des Prüfungsraums eingerichtet. Dort können an einem JKU-Servicepoint die Funktion der JKU Card (vgl. Kapitel 2) kontrolliert und ihre Gültigkeit verlängert, sowie an einem PC mit Chipkartenleser die Funktionstüchtigkeit der Bürgerkarte überprüft werden. Die benötigte Serverinfrastruktur mit unterbrechungsfreier Stromversorgung ist in den gleichen Räumlichkeiten untergebracht, sodass die meisten Prüfungen völlig autark durchgeführt werden können. Lediglich für die Berechtigungsprüfung, die digitale Signatur und für Prüfungen mit freigegebenem Internetzugang wird Netzwerkkonnektivität zur Universität benötigt.

Abb. 1 zeigt die eigens im Haus entwickelte Benutzeroberfläche des im Prüfungsraum befindlichen Steuerungs-PCs. Mit einem Klick auf das „Play“-Symbol kann ein einzelner PC im aktuell ausgewählten Prüfungsmodus (hier „Moodle-Prüfung“) eingeschaltet werden; mit den Symbolen links unten funktioniert das auch für mehrere oder alle gleichzeitig. Eingeschaltete PCs zeigen das in der zentralen Datenbank abgespeicherte Foto der Studierenden und ihre Matrikelnummer. Zudem ist es hier möglich, einen Live-Stream des Bildschirms anzusehen, um den Stand der Prüfung mitzuverfolgen oder verdächtige Aktivitäten besser beurteilen zu können. Weiters kann der PC zurückgesetzt werden, um bspw. schnell den Prüfungsmodus wechseln zu können oder um bei technischen Problemen einen Neustart durchzuführen. Im rechten unteren Teil der Oberfläche sind Steuerungs- und Auswertungsmöglichkeiten untergebracht. Hier kann der Prüfungsmodus gewählt werden – bei manchen Prüfungen ist eine vollständige Arbeitsumgebung notwendig, um etwa Aufgaben in Office-Programmen zu erledigen, während bei anderen Prüfungen lediglich ein Browser zum Zugriff auf eines der Prüfungsmoodle benötigt wird. Für spezielle Prüfungen wurden eigene Moodle-

Instanzen eingerichtet. Zudem kann ausgewählt werden, ob ein voller oder eingeschränkter Internetzugang möglich sein soll, diese Auswahl wird dann per Script in der Firewall entsprechend umgesetzt und ist sofort wirksam. Außerdem können hier verschiedene Auswertungen erzeugt und Informationen abgerufen werden.

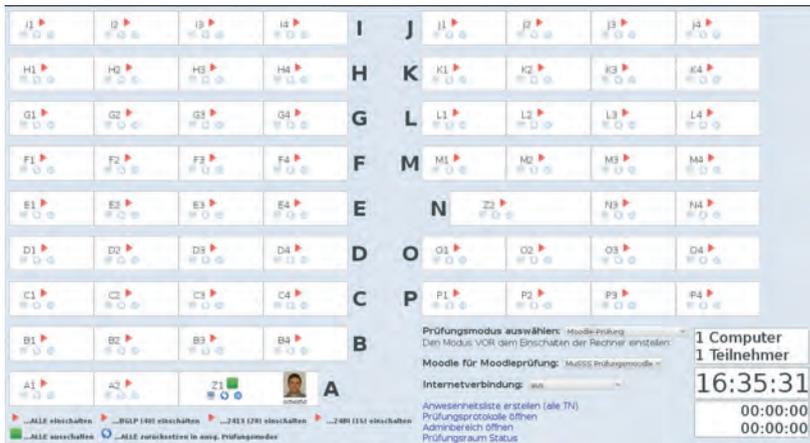


Abb. 1: Steuerungs-PC

Zur Prüfung melden sich die Studierenden an einem der PCs an, in dem sie ihre JKU Card in das in die Tastatur integrierte Lesegerät stecken. Daraufhin wird mit der Webcam ein Foto der Person gemacht, die vor dem PC sitzt. Dieses kann von der Prüfungsaufsicht mit dem in der zentralen Datenbank gespeicherten Foto verglichen werden. So kann sichergestellt werden, dass es sich bei allen Prüflingen um jene Personen handelt, die sie vorgeben zu sein. Nach dem Anmeldevorgang landen die Studierenden in der festgelegten Arbeitsumgebung. Bei einer Prüfung in Moodle kommt bereits vor Beginn der Prüfung die digitale Signatur zum Einsatz: Die Studierenden bestätigen die gültigen Richtlinien über die Durchführung von Computerprüfungen⁴.

Nach Abschluss der Prüfung ist eine weitere digitale Signatur zu leisten. Dazu werden die gegebenen Antworten nochmals angezeigt und die Studierenden bestätigen, dass es sich tatsächlich um die von ihnen gegebenen Antworten handelt. Im Anschluss an diese zweite Signatur werden bei den meisten Prüfungen unmittelbar Auswertung und Ergebnisse angezeigt, dies zählt auch gleichzeitig als Einsichtnahme, da zumeist auch Hinweise zu den richtigen Lösungen angezeigt werden. Nach Beendigung der Prüfung wird eine automatisch generierte Anwesenheitsliste gedruckt und per E-Mail versendet. Ebenso wird ein Prüfungsprotokoll mit Start- und Endzeitpunkt der Prüfung und ähnlichen Informationen generiert, welches um manuelle Einträge ergänzt werden kann. Sollte jemand während der Prüfung den Prüfungsraum verlassen, wird die JKU Card in einen vorgesehenen Kartenleser in der Nähe der Ausgangstür gesteckt und beim

⁴ Zu finden unter <http://www.jku.at/mtb?jahr=2014&nr=139>

Zurückkehren in den Raum wieder mitgenommen. Dadurch wird ein entsprechender Eintrag im Prüfungsprotokoll erzeugt. Ein weiterer Sicherheitsmechanismus sind Screenshots, die alle 5 Sekunden erzeugt und zentral abgespeichert werden, so den Fortschritt dokumentieren und bei eventuellen Problemen oder Ungereimtheiten der Nachvollziehbarkeit dienen.

4 Erfahrungen

Seit Sommersemester 2011 wird der Prüfungsraum verwendet, in diesem Jahr fanden 54 Prüfungsdurchläufe mit insgesamt 1267 Studierenden statt, angemeldet waren 1583, geprüft wurde in 7 Fächern bzw. Modulen. Bis 2016 hat sich die Zahl der Prüfungsdurchläufe knapp verdreifacht und die geprüften Studierenden mehr als vervierfacht. Die Anzahl der geprüften Module stieg auf 23, der Anstieg ist u. a. durch den Start der Mediziner-Ausbildung im Herbst 2016 bedingt. Die positive Empfehlung von Kolleginnen und Kollegen hat sich auch hier als effiziente Werbemaßnahme bestätigt. Die digitale Signatur wird seit WS 2012/13 für die Prüfungen verwendet, in den ersten drei Semestern übernahm das Zulassungsservice die Ausstellung der Bürgerkarten. Seit SS 2014 erfolgt die Ausstellung im Rahmen der Administration des Prüfungsraumes. Seit SS 2014 wurden knapp 5500 digitale Signaturen ausgestellt.

Bei den Prüfungen am Ende des Wintersemesters 2016/17 wurde mittels eines Stimmungsbarometers Rückmeldungen der Studierenden zur Prüfungsorganisation und zum Prüfungsraum im Allgemeinen eingeholt. Bei elf Prüfungen mit 536 Studierenden wurden 399 Fragebögen abgegeben, was einem Rücklauf von 75% entspricht, 33 beantworteten nur die Fragen der ersten Seite. 53% belegen ein Bachelorstudium, 43% sind im Master und 4% belegen ein Doktoratsstudium, der Frauenanteil beträgt 64%, nur 28% der Studierenden sind nicht berufstätig, 19% sind voll berufstätig. Die Studierenden bewerten die Organisation der Prüfung, die Arbeitsumgebung und den Prüfungsablauf auf einer 5-teiligen Likertskala mit Schulnoten. Die Prüfungsorganisation wird durchwegs positiv bewertet, die Durchschnittswerte bewegen sich zwischen 1,4 und 1,8. Die Doktoratsstudierenden werten durchwegs besser als die Bachelorstudierenden mit Ausnahme der Ausstellung der Bürgerkarte. Die Mittelwerte der Bewertungen der Arbeitsumgebung liegen zwischen 1,7 und 2,2 auf der fünfteiligen Skala. Vor allem die vorgebeugte Sitzposition und die Bildschirme werden nicht so gut bewertet. Die Prüfungsabhaltung wird sehr positiv bewertet, vor allem die unmittelbare Prüfungseinsicht und Verfügbarkeit der Ergebnisse (Ausnahme Doktoratsstudierende, bei diesem Prüfungsdurchlauf wurden die Ergebnisse erst später bekannt gegeben). Die Verwendung der digitalen Signatur wird von den Studierenden äußerst positiv bewertet.

5 Fazit und Ausblick

Die in dieser Forschung präsentierten Zahlen zeigen, dass der elektronische Prüfungs-

raum sowohl von Lehrenden als auch von Studierenden sehr gut angenommen wird. Er hat sich in relativ kurzer Zeit als fixer Bestandteil der universitären Prüfungsumgebung etabliert und wird regelmäßig genutzt. Auch die Technik arbeitet inzwischen äußerst stabil, es kommt nur extrem selten zu Problemen, meist sind diese auf Bedienungsfehler oder defekte Hardware zurückzuführen. Die flächendeckende Ausrollung der Bürgerkarten auf den JKU-Cards schreitet gut voran.

In Rottenmann, einem der Außenstandorte der JKU, wurde nach ähnlichem Muster bereits ein weiterer, mit 12 Plätzen ausgestatteter Prüfungsraum eingerichtet. Prüfungen können dort entweder autonom oder auch gleichzeitig mit Prüfungen im Hauptprüfungsraum stattfinden. Die JKU wurde 2014 um eine medizinische Fakultät erweitert, auch diese nutzt den Prüfungsraum bereits intensiv und prüft gerade die Einrichtung eines eigenen Prüfungsraums in einem der Neubauten.

Literaturverzeichnis

- [BEH09] Biella, D.; Engert, S.; Huth, D.: Design and delivery of an e-assessment solution at the University of Duisburg-Essen. In: Proceedings of the European University Information Systems Organisation Conference (EUNIS 2009) https://www.uni-due.de/imperia/md/content/zim/veranstaltungen/eunis_09_e-assessment.pdf, Stand: 08.03.2017.
- [Fi13] Fischer, H.: E-Learning im Lehralltag. Analyse der Adoption von E-Learning in der Hochschullehre. Springer, VS-Verlag, Wiesbaden, 2013.
- [Ha16] Halbherr, T.; Dittmann-Domenichini, N.; Piendl, T.; Schlienger, C.: Authentische, kompetenzorientierte Online-Prüfungen an der ETH Zürich, Zeitschrift für Hochschulentwicklung, S. 247-269, 2016.
- [Hu10] Huth, D.: E-Assessment an der Universität Duisburg-Essen im „Kompetenzzentrum für PC-gestützte Prüfungen“, [Präsentationsfolien, 2. Symposium E-Learning an Hochschulen, 9. 3. 2010], https://www.uni-due.de/imperia/md/content/zim/projekte/e-assessment_ude_09032010_v2.pdf, Stand: 08.03.2017.
- [KB16] Karami, M.; Baumann, M.: Zum Glück gezwungen? Wie die Einführung eines hochschulweiten Angebots elektronischer Prüfungen gelingen kann. In (Brahm, T.; Jenert, T.; Euler, D., Hrsg.): Pädagogische Hochschulentwicklung: Von der Programmatik zur Implementierung. Wiesbaden: Springer Fachmedien Wiesbaden, S. 313-323, 2016.
- [Ko16] Kopp, M.; Ebner, M.; Rehatschek, H.; Schweighofer, P.; Sfiri, A.; Swoboda, B.; Teufel, M.: Entwicklung von „eDidactics“: Ein Fortbildungsprogramm für den Einsatz von Technologien in der Hochschullehre, Qualität in Studium und Lehre: Kompetenz- und Wissensmanagement im steirischen Hochschulraum. Wiesbaden, Springer Fachmedien Wiesbaden, S. 305-321, 2016.
- [Lo16] Loviscach, J.: Digitalisierung der Hochschulausbildung: zwischen Graswurzel und Big Business, 2016, www.j3L7h.de, Stand: 03.02.2017.

Automatische Sammlung von Aktivitäten Lernender in offenen Lernumgebungen und deren Nutzung in einer Lerntagebuchanwendung

Svenja Neitzel¹ und Christoph Rensing²

Abstract: Lernen erfolgt heute zunehmend in offenen Lernumgebungen, die u.a. dadurch charakterisiert sind, dass verschiedene Lernanwendungen von Lernenden individuell verwendet werden. In diesem Beitrag werden eine Architektur und deren verschiedene Komponenten vorgestellt, die es erlauben, in solchen offenen Lernumgebungen Informationen über Lernaktivitäten zu sammeln und auszuwerten. Als Motivation und Beispiel für die Verwendung dieser Informationen dient eine eigens entwickelte Lerntagebuchanwendung. Die Potenziale des Einsatzes von Lerntagebüchern zur Planung, Dokumentation und Reflexion des eigenen Lernverhaltens im selbstgesteuerten Lernen sind zwischenzeitlich anerkannt. Allerdings bedeutet die Pflege eines detaillierten Lerntagebuchs für den Lernenden einen relativ hohen Aufwand. Erfolgt die Pflege der Lerntagebucheinträge nicht zeitnah, besteht die Gefahr, dass einzelne Einträge vergessen werden oder von der Realität abweichen. Mittels der beschriebenen Architektur lassen sich Lerntagebucheinträge teilweise automatisch erstellen. Außerdem erlaubt sie einen Vergleich des eigenen Lernverhaltens mit dem Lernverhalten anderer Lernender. Gleichzeitig sichert unsere Architektur die Privatheit der Daten mittels der Verwendung eines plattformübergreifend genutzten Pseudonyms.

Keywords: Lerntagebuch, Lernanwendungen, Learning Analytics, Selbstgesteuertes Lernen, Reflexion.

1 Motivation

Unter anderem begünstigt durch den technologischen Fortschritt findet Lernen heute nicht mehr nur in vorgegebenen formellen Strukturen, sondern zunehmend selbstgesteuert statt. Dies gilt auch für das Lernen an Hochschulen: Studierende besuchen nicht nur Präsenzveranstaltungen, sondern können mit Hilfe von Online-Angeboten - wie Vorlesungsaufzeichnungen - zunehmend selbst entscheiden, wann und wo sie lernen. Eine Vielzahl von technischen Hilfsmitteln wie Foren, E-Mail oder Instant Messaging steht darüber hinaus zur Verfügung, um unabhängig von Ort und Zeit mit anderen Lernenden zu kooperieren. Weiterhin bieten verschiedene Websites und Apps eine Vielzahl an Diensten und Ressourcen, die von Lernenden, neben den vom Lehrenden zur Verfügung gestellten Materialien, genutzt werden können. Wir sprechen dabei von

¹ Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstr. 10, 64283 Darmstadt, svenja.neitzel@kom.tu-darmstadt.de

² Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstr. 10, 64283 Darmstadt, christoph.rensing@kom.tu-darmstadt.de

offenen Lernumgebungen. Bei der Nutzung der verschiedenen Lernanwendungen produzieren Lernende Daten, die analysiert werden können um Lernverhalten besser zu verstehen sowie Adaptions- und Interventionsmechanismen anzubieten. In offenen Lernumgebungen repräsentieren die Daten einzelner Anwendungen jedoch nur Ausschnitte des gesamten Lernverhaltens. Eine zentrale Sammlung und Auswertung von Daten mehrerer Anwendungen ist darum wünschenswert, wirft aber die Frage auf, wie dabei die Privatheit der Daten gewährleistet werden kann.

Das oben beschriebene Szenario entspricht in weiten Teilen der Definition von selbstgesteuertem Lernen: "Selbstgesteuert ist ein Lernen dann, wenn die Operationen/Strategien, Kontrollprozesse und der Offenheitsgrad vom Subjekt auf ein Ziel hin selbstbestimmt werden" [Ki04, S. 85]. Die Vorteile des selbstgesteuerten Lernens gehen einher mit Herausforderungen. So verlangt selbstgesteuertes Lernen gewisse Selbststeuerungskompetenzen. Viele Lernende haben jedoch Schwierigkeiten, ihr Lernen selbst zu organisieren sowie Motivation aufzubauen und aufrecht zu erhalten. Der Einsatz von Lerntagebüchern, in denen die Lernenden ihre Lernaktivitäten planen, dokumentieren und reflektieren, bietet Potenziale zur Verbesserung der Organisation des Lernprozesses und Aufrechterhaltung der Motivation. Die Pflege der Einträge in einem Lerntagebuch ist jedoch mitunter aufwändig, insbesondere wenn es sich nicht nur um wiederkehrende, sondern vielfältige Lernaktivitäten handelt und diese oftmals eher kurze Zeiträume umfassen. Dies ist beim selbstgesteuerten Lernen zunehmend zu beobachten.

Die im vorliegenden Beitrag betrachtete primäre Forschungsfrage ist, ob und wie Informationen über Lernaktivitäten aus verschiedenen Lernanwendungen gesammelt werden können, sodass die Privatheit weitgehend gewährleistet ist und zugleich Analysen der Lernaktivitäten erfolgen können. Eine entsprechende Architektur wird vorgestellt und eine beispielhafte Lerntagebuchanwendung beschrieben, welcher es die Architektur ermöglicht, Einträge über Lernaktivitäten teilweise automatisch zu erfassen und Vergleichsdaten anderer Lernender zu beziehen. Weiterhin beschäftigt sich der Beitrag mit der Gestaltung der Lerntagebuchanwendung selbst. In unserer Betrachtung stehen dabei die metakognitiven Aspekte der Selbststeuerung im Fokus, weshalb wir uns auf die zeitliche Planung und Dokumentation von unterschiedlichen Lernaktivitäten sowie Möglichkeiten zur Reflexion beschränken. Eine Unterstützung der inhaltlichen Reflexion des Gelernten auf kognitiver Ebene haben wir nicht zum Ziel.

Zunächst vertiefen wir nachfolgend die Motivation, indem wir verwandte Arbeiten zur Unterstützung der Selbststeuerung im Lernen vorstellen. Zudem betrachten wir die Themenstellung der Sammlung und Auswertung von Daten über Lernaktivitäten aus dem Blickwinkel des Forschungsbereichs *Learning Analytics*. Im Abschnitt 3 stellen wir das Konzept zur Erreichung der Zielsetzung vor und in Abschnitt 4 dessen Umsetzung. Der Beitrag schließt mit einer Zusammenfassung und mit einem Ausblick auf weitere Arbeiten.

2 Verwandte Arbeiten

Selbstgesteuertes Lernen und die dieser Form des Lernens zugrundeliegenden Verarbeitungsprozesse wurden in der Psychologie ausführlich untersucht. Beispielsweise unterscheidet Boekarts [Bo99] drei Systeme (das *kognitive System*, das *motivationale System* und das *metakognitive System*), die Lernende steuern müssen. Die Kompetenz „selbstgesteuert zu lernen“ als Teil des metakognitiven Systems muss explizit erworben werden. In [SW06] werden dazu drei Phasen („vor dem Lernen“, „während des Lernens“ und „nach dem Lernen“) unterschieden. In der Phase vor dem Lernen sollen Lernziele gesetzt werden und eine Lernstrategie definiert werden. In der Phase nach dem Lernen sollen eine Reflexion und gegebenenfalls eine Anpassung der Ziele und Strategien erfolgen. Die Nutzung eines Lerntagebuches, also eines selbstgeführten Protokolls über den eigenen Lernprozess, zur Planung und zur späteren Reflexion, hat sich in verschiedenen Untersuchungen als erfolgreiches Mittel zur effektiveren Selbststeuerung herausgestellt [GH07]. Auch in [RN02] wird die Bedeutung des Lerntagebuches für den Erwerb von Selbststeuerungskompetenzen auf der metakognitiven Ebene betont.

Es gibt verschiedene digitale Werkzeuge, um Planung und Reflexion im selbstgesteuerten Lernen zu unterstützen. Erfolgt eine solche Unterstützung, ergibt sich ein hohes Potenzial zur Verbesserung der Lernqualität [Be10]. Beispielhaft sei ein Werkzeug zur Vermittlung von Selbststeuerungskompetenzen genannt [Sc09], welches in eine Lernanwendung integriert ist und es den Lernenden erlaubt, ihre Ziele für eine selbstgesteuerte Lernphase explizit zu machen sowie aktiv zur Reflexion des eigenen Verhaltens auffordert. [Ko16] erlauben mit ihrem Werkzeug eine individuelle Planung der nächsten Lernphase und nachfolgend einen Abgleich zwischen Planung und tatsächlichem Lernverhalten sowie den Vergleich des eigenen Lernverhaltens mit dem einer Gruppe von Peers. Die Herausforderung der Definition einer für den Vergleich geeigneten Gruppe von Peers diskutiert [Wi14].

Die Sammlung und Auswertung von Daten über Lernaktivitäten ist insgesamt Gegenstand des Forschungsfeldes *Learning Analytics*. Eine Zielsetzung von Learning Analytics ist die Bereitstellung von Informationen über Lernaktivitäten zum Zwecke der Reflexion und der darauf basierenden Optimierung des Lernverhaltens [Dy13]. Diese Zielsetzung stimmt mit der unseres Vorhabens überein. Der überwiegende Anteil der Arbeiten im Bereich Learning Analytics beschränkt sich jedoch auf die Sammlung und Auswertung von Daten über Lernaktivitäten, die in einem geschlossenen System, wie einer speziellen Lernanwendung oder einem Learning-Management-System (LMS), anfallen. Dies steht im Widerspruch zu unserer Zielsetzung, Daten aus verschiedenen von Lernenden genutzten Anwendungen zu sammeln und auszuwerten. Die *xAPI* Spezifikation und die Entwicklung von sogenannten *Learning Record Stores* (LRSs) sind weitgehend akzeptierte Ansätze, um dieser Herausforderung zu begegnen. Eine große Anzahl von Wissenschaftlern und Entwicklern folgen dem Konzept einer zentralisierten Speicherung der Daten und nutzen zur Sammlung der Daten die *xAPI* Spezifikation [KR16]. *xAPI* definiert sogenannte *Activity Statements*, bestehend aus

Aktor, Verb und Objekt. Der Akteur beschreibt den Lernenden, der eine Lernaktivität durchführt. Das Verb und Objekt charakterisieren die durchgeführte Aktivität sowie das Objekt mit dem die Aktivität durchgeführt wurde.

3 Konzept

Entsprechend der Zielsetzung unserer Arbeit und der zuvor formulierten Forschungsfragen, sind Konzepte in zwei Bereichen zu entwickeln: Einerseits wird ein Konzept für die Lerntagebuchanwendung selbst benötigt. Andererseits sind Konzepte zur Sammlung sowie zur Auswertung und Nutzung von Daten über Lernaktivitäten aus verschiedenen Lernanwendungen notwendig. Bevor nachfolgend diese Konzepte behandelt werden, wird zunächst das betrachtete Lernszenario genauer beschrieben und die mittels der Gesamtanwendung verfolgten Zielsetzungen für verschiedene Beteiligte dargestellt.

3.1 Lernszenario

Das betrachtete Lernszenario, vgl. Abbildung 1, ist vor allem im Hochschulbereich häufig anzutreffen: Der Lehrende nutzt das LMS seiner Bildungsinstitution, um seinen Lernenden (Kohorte) verschiedene Dienste und Inhalte zur Verfügung zu stellen. Zusätzlich nutzen die Lernenden individuell gewählte Lernanwendungen. Diese liegen außerhalb der Einflussphäre der Bildungsinstitution. In unserer Arbeit ergänzen wir das Szenario um eine Lerntagebuchanwendung, welche alle Lernenden nutzen können.

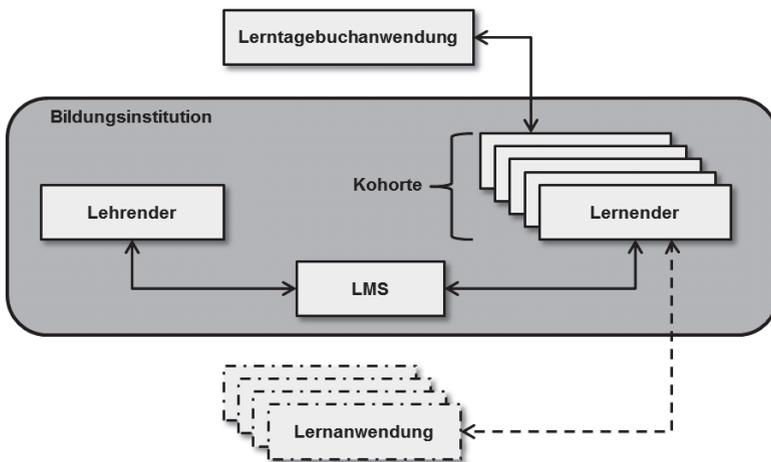


Abb. 1: Betrachtetes Lernszenario einer offenen Lernumgebung ergänzt um eine Lerntagebuchanwendung

3.2 Ziele der Gesamtanwendung

Die Zielsetzung von Learning Analytics wird oftmals für drei Stakeholder (Lernende, Lehrende und Bildungsinstitutionen bzw. Bildungsforschung) unterschieden. Wir folgen dieser Unterscheidung und konkretisieren die Zielsetzungen anhand des betrachteten Lernszenarios:

- Ziele für Lernende: Das für Lernende verfolgte Ziel ist die Bereitstellung von Informationen zur Reflexion des eigenen Lernverhaltens und zum Vergleich mit dem Lernverhalten anderer Lernender aus der gleichen Kohorte [Dy13]. Insbesondere der Vergleich mit anderen bietet Referenzwerte, um das eigene Verhalten beurteilen zu können [Wi14]. Durch eine zentrale Datensammlung soll im betrachteten Szenario ein Vergleich mit anderen Lernenden ermöglicht werden. Die Lerntagebuchanwendung stellt die Daten dazu bereit. Zusätzlich soll die Lerntagebuchanwendung eine Reflexion von geplantem und tatsächlichem Lernverhalten ermöglichen, um ggf. Verhaltensänderungen vorzunehmen.
- Ziele für Lehrende: Lehrende sind nach [Dy13] eine weitere Zielgruppe von Learning Analytics. Für die Beurteilung des eigenen Lehrangebotes kann es sehr hilfreich sein, Informationen über die Nutzung der angebotenen Dienste und Inhalte zu erhalten. Dies ist für selbstgesteuerte Lernformen deutlich wichtiger als für traditionelle Lernformen in Präsenz. Im vorliegenden Szenario kann der Lernende durch die Lerntagebuchanwendung detaillierte Informationen über sein Lernverhalten bereitstellen, auf die der Lehrende durch die zentrale Datensammlung Zugang erhalten kann.
- Ziele für die Bildungsforschung: Ziel von Learning Analytics ist auch die Gewinnung von Erkenntnissen in forschungsrelevanten Bereichen. Dazu gehört beispielsweise die Ermittlung von Zusammenhängen zwischen Lernverhalten und erzielten Ergebnissen, was z.B. in [SM11] umfangreich untersucht wurde. Die Kenntnis über Zusammenhänge zwischen Lernverhalten und erzielten Ergebnissen kann der Vorhersage des Lernerfolgs dienen [Ro13] und damit Grundlage für Interventionen sein. Die Lerntagebuchanwendung kann interessante Daten zur Erforschung solcher Zusammenhänge bereitstellen und ist auch eine potenziell geeignete Anwendung um zukünftig Interventionen anzubieten, die dann wiederum den Lernenden zugutekommen.

3.3 Konzept der Lerntagebuchanwendung

Entsprechend der Zielsetzungen, soll die Lerntagebuchanwendung folgende Funktionen anbieten:

- Planung von Lernaktivitäten,

- Automatische Erfassung von Lernaktivitäten aus anderen Lernanwendungen und manuelle Erfassung von durchgeführten Lernaktivitäten für nicht automatisch erfassbare Lernaktivitäten,
- Visualisierung des eigenen Lernverhaltens mit Berücksichtigung des Status der Lernaktivitäten (geplant oder durchgeführt) zum Zwecke der Reflexion,
- Visualisierung der eigenen durchgeführten Lernaktivitäten im Vergleich zu den durchgeführten Lernaktivitäten einer Peer-Gruppe,
- Erinnerungsfunktionen zur regelmäßigen Pflege des Lerntagebuches um die Vollständigkeit und Korrektheit der Daten zu erhöhen.

Lernaktivitäten sind über folgende Attribute näher beschrieben:

- Modul: Das Modul gibt an welchem Kurs die Lernaktivität zugeordnet wird.
- Lernaktivitätstyp: Typen von Aktivitäten sind beispielsweise ein Vorlesungsbesuch, die Bearbeitung einer Übung oder das Lernen in einer Gruppe.
- Zeitliche Angaben: Diese Angaben beinhalten mindestens Tag und Dauer der Lernaktivität, können aber auch Zeitpunkte für Beginn und Ende umfassen.
- Status: Der Status einer Lernaktivität kann „geplant“ oder „durchgeführt“ sein.

3.4 Konzept zu Datensammlung, Identitätsmanagement und Datenanalyse

Die kombinierte Nutzung von Lernaktivitäten aus verschiedenen Lernanwendungen setzt einen Datenaustausch mit anderen Lernanwendungen voraus. Anstelle eines direkten Datenaustausches zwischen den einzelnen Anwendungen folgen wir dem zuvor vorgestellten Konzept eines zentralen LRS. Dies bedeutet, dass die von Lernenden genutzten Lernanwendungen protokollierte Lernaktivitäten an den LRS übertragen. Andere Anwendungen, wie in unserem Szenario die Lerntagebuchanwendung, können daraufhin alle Lernaktivitäten eines Lernenden vom LRS abrufen.

In einer offenen Lernumgebung ergibt sich dabei die Herausforderung der korrekten Identifizierung einzelner Lernender: Da Lernende in verschiedenen Lernanwendungen häufig mit verschiedenen Identitäten registriert sind, bedarf es eines Konzeptes zum Identitätsmanagement. Zugleich soll jedoch die Privatheit der Daten sichergestellt werden. Zur Realisierung dieser beiden Anforderungen verwenden wir den sogenannten *Pseudonymity Provider* [Rö17] als eigene Infrastrukturkomponente.

Der Vergleich des eigenen Lernverhaltens mit dem Lernverhalten anderer Lernender verlangt einen Zugriff auf deren Daten im LRS. Diese können von dort in pseudonymisierter Form abgerufen werden. Da die Datenmengen in großen Gruppen sehr umfangreich werden können und mehrere Nutzer identische Daten benötigen, erfolgt eine Aggregation der Daten bereits durch entsprechend formulierte Abfragen an den LRS. Somit kann die Menge der zu übertragenden Daten reduziert werden.

4 Implementierung

Das zuvor beschriebene Konzept wurde unter Verwendung der in Abbildung 2 dargestellten Komponenten implementiert. Als LMS dient die Open Source Lernplattform *Moodle*³. Das ebenfalls als Open Source Produkt zur Verfügung stehende *Learning Locker*⁴ wird als LRS genutzt. Learning Locker verwendet xAPI und bietet neben der Speicherung der Activity Statements auch ein einfaches Dashboard mit der Möglichkeit, Analysen hinsichtlich verschiedener Aspekte zu erstellen. Dies ermöglicht dem Lehrenden bereits grundlegende Einblicke in das Lernverhalten seiner Kursteilnehmer. Zum Identitätsmanagement wird der in [Rö17] beschriebene Pseudonymity Provider verwendet. Bei der Lerntagebuchanwendung handelt es sich um eigene Entwicklungen, welche nachfolgend detailliert beschrieben wird. Zunächst wird jedoch auf die Verwendung des Pseudonymity Providers im betrachteten Szenario eingegangen.

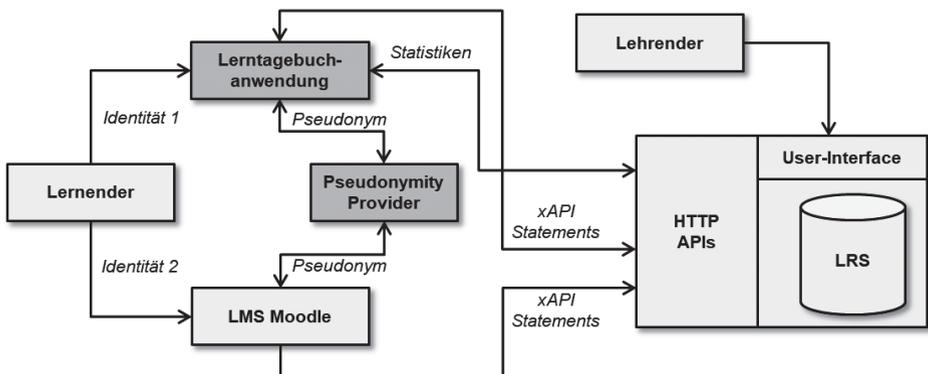


Abb. 2: Gesamtarchitektur

4.1 Identitätsmanagement und Wahrung der Privatheit

Die Aufgabe des Pseudonymity Providers besteht in der Generierung eines eindeutigen Pseudonyms für einen Lernenden und die Bereitstellung dieses Pseudonyms an verschiedene Lernanwendungen, unabhängig davon, ob der Lernende in diesen Lernanwendungen verschiedene Identitäten verwendet. Der Lernende muss sich dazu einmalig einen Account beim Pseudonymity Provider erstellen, wodurch dieser ein eindeutiges Pseudonym für ihn generiert. Die Lernanwendungen müssen sich ebenfalls beim Pseudonymity Provider registrieren, um Verfahren und Parameter zur sicheren Datenübertragung zu spezifizieren. Weiterhin müssen die Lernanwendungen zur

³ <https://moodle.de/>

⁴ <https://learninglocker.net/>

Verwendung des Pseudonyms geringfügig erweitert werden: Sie müssen zum einen eine Möglichkeit bieten, dass der Lernende der Anwendung die Zustimmung zur Abfrage des Pseudonyms beim Pseudonymity Provider und zur Verwendung des Pseudonyms gibt. Zum anderen müssen die Lernanwendungen das Pseudonym anstelle von der auf dieser Plattform verwendeten Identität des Lernenden innerhalb der xAPI Statements als Aktor nutzen. In Moodle verwenden wir dazu das ebenfalls in [Rö17] beschriebene *PseudoLearner* Plug-In. Aus diesem wird der Lernende an den Pseudonymity Provider weitergeleitet, gibt dort seine Zugangsdaten ein und stimmt der Weitergabe des Pseudonyms an diese Moodle-Instanz zu. Anschließend kann der Lernende für jeden seiner Kurse die Zustimmung zur Nutzung des Pseudonyms erteilen oder entziehen. Weiterhin wurde das *Logstore_xAPI* Plug-In⁵ für Moodle so modifiziert, dass es, im Falle eines vorhandenen Pseudonyms, dieses anstelle des in Moodle hinterlegten Benutzernamens für die xAPI Statements verwendet. In der Lerntagebuchanwendung, welche als Android App implementiert wurde, verwenden wir ein eigenes implementiertes Modul welches ebenfalls die Möglichkeit zur Eingabe der Zugangsdaten zum Pseudonymity Provider bereitstellt, mit dem Pseudonymity Provider kommuniziert und schließlich das Pseudonym an die App zurückliefert. Dieses Modul ist unabhängig von der Lerntagebuchanwendung und somit ohne weiteres auch von anderen Android Apps nutzbar.

4.2 Lerntagebuchanwendung

Die Lerntagebuchanwendung wurde als Smartphone App für das mobile Betriebssystem Android implementiert und bietet im Wesentlichen die in Kapitel 3.3 vorgestellten Funktionen. Für die Pflege des Lerntagebuchs ist folgendes Vorgehen vorgesehen: Die grobe Planung von Lernaktivitäten für die kommende Woche soll am Ende der aktuellen Woche erfolgen. Dazu kann der Lernende Lernaktivitäten „grob“ (d.h. nur mit Tag und Dauer) planen. Zur Planung der Lernaktivität muss neben den zeitlichen Angaben ein Lernmodul und einen Lernaktivitätstyp ausgewählt werden⁶. An Tagen der kommenden Woche, an denen eine Lernaktivität geplant wurde, soll der Lernende diese morgens „konkret“ planen (d.h. eine Uhrzeit definieren). Am Abend sollen die geplanten Lernaktivitäten als erledigt markiert oder ggf. geändert werden. Der Lernende wird durch Erinnerungen in Form von Benachrichtigungen an dieses regelmäßige Vorgehen herangeführt. Die Lerntagebuchanwendung kann jedoch auch völlig anders genutzt werden, da alle Funktionen jederzeit zur Verfügung stehen. Abgesehen von der Erinnerung zur groben Planung der kommenden Woche erfolgen die Benachrichtigungen adaptiv, d.h. es werden nur Benachrichtigungen zur Bearbeitung geplanter Lernaktivitäten angezeigt, wenn geplante Aktivitäten vorliegen oder aber Aktivitäten aus anderen Lernanwendungen automatisch erfasst wurden und keiner im Lerntagebuch angelegten Lernaktivität zugeordnet werden konnten.

⁵ https://moodle.org/plugins/logstore_xapi

⁶ Für die erste Version der Lerntagebuchanwendung gehen wir dafür von einem vordefinierten Modul und einer ebenfalls vorgegebenen Menge von Lernaktivitätstypen aus.

Die Erfassung von Aktivitäten aus anderen Lernanwendungen durch die Lerntagebuchanwendung ist möglich, sofern die Anwendungen protokollierte Lernaktivitäten unter dem plattformübergreifenden Pseudonym an den gemeinsamen LRS senden. In regelmäßigen Abständen fordert die Lerntagebuchanwendung die neuen Lernaktivitäten des Lernenden vom LRS an und speichert diese lokal. Die Lerntagebuchanwendung versucht zudem, erfasste Lernaktivitäten anderer Lernanwendungen, einer im Lerntagebuch geplanten Lernaktivität zuzuordnen. Dies geschieht momentan auf rein zeitlicher Basis (d.h. befindet sich eine erfasste Lernaktivität zeitlich innerhalb einer im Lerntagebuch vorhandenen, werden diese als zusammengehörig interpretiert). Dieser interessante Aspekt der Zuordnung verschiedener Lernaktivitäten bietet vor allem in komplexeren Szenarien noch viel Raum für Verbesserungen und könnte auch dazu verwendet werden, automatisch zu erkennen ob geplante Lernaktivitäten tatsächlich durchgeführt wurden.

Zur Gewinnung von Forschungsdaten und um langfristig die Funktionalität der Lerntagebuchanwendung zu erweitern, werden noch weitere Informationen durch die Lerntagebuchanwendung erfasst. So wird zum Ende der Woche ein Fragebogen mit verschiedenen Items zur Reflexion des Lernverhaltens in der vergangenen Woche angeboten. Zudem sammelt die Lerntagebuchanwendung, sofern die Lernenden dem zustimmen, während der Durchführung von Lernaktivitäten kurze anonyme Audio-Snippets. Zusammen mit anderen Sensordaten sollen diese zukünftig zum Training von Klassifikatoren verwendet werden, die ebenfalls zur automatischen Erkennung von Lernaktivitäten, wie in [DRT14] beschrieben, verwendet werden können.

Zur Reflexion des Lernverhaltens sind in der App verschiedene Visualisierungen der durchgeführten Lernaktivitäten mit Berücksichtigung der Abweichung zwischen geplanten und tatsächlich durchgeführten Lernaktivitäten realisiert. Abbildung 3 zeigt beispielhaft eine Tages-, Wochen- und Semesteransicht der eingetragenen Lernaktivitäten. In der Tagesansicht werden die für diesen Tag geplanten und durchgeführten Lernaktivitäten aufgeführt und deren Verhältnis visualisiert. In dieser Ansicht ist auch die Zuordnung von automatisch erfassten Lernaktivitäten sichtbar. Im Falle von Moodle haben erfasste Activity Statements keine Dauer und werden daher mit Null Minuten angezeigt und ausgewertet. Dies liegt darin begründet, dass Moodle nur einzelne Klicks als Activity Statements in den LRS schreibt. Informationen darüber wie lange sich der Lernende tatsächlich mit dem angeklickten Inhalt beschäftigt hat, liegen in Moodle und damit auch im LRS nicht vor. Dennoch haben auch diese Lernaktivitäten einen großen Mehrwert für die Lerntagebuchanwendung: Beispielsweise kann der Lernende anhand seiner Aktivitäten in Moodle genau erkennen, wann bestimmte Ressourcen von ihm genutzt wurden und erhält dadurch Anhaltspunkte für das realitätsgetreue Dokumentieren seiner Lernaktivitäten im Lerntagebuch. Wie oben bereits beschrieben, erhält auch die Lerntagebuchanwendung selbst durch die Aktivitäten aus Moodle Hinweise darüber, dass der Lernende in einer bestimmten Zeit aktiv war und adaptiert ihre Funktionen entsprechend. In der Wochenansicht plant der Lernende seine Lernaktivitäten mit Tag und Dauer für die kommende Woche. Das tatsächliche Lernverhalten im Laufe der Woche kann hier mit dem geplanten Lernverhalten verglichen werden, wobei geplante

Lernaktivitäten durch eine größere Farbtransparenz und eine gestrichelte Linienform visuell von durchgeführten Lernaktivitäten unterschieden werden. In der Semesteransicht sind im oberen Bereich die summierten Zeitaufwände pro Semesterwoche visualisiert. Die gestrichelte und leicht transparente Linie gibt hier den durchschnittlichen Zeitaufwand der Peer-Gruppe an, wobei die Peer-Gruppe hier alle Lernenden des Moduls umfasst. Die Lerntagebuchanwendung fragt diese Werte beim LRS an, speichert sie und aktualisiert sie in regelmäßigen Abständen. Der verwendete LRS bietet hierfür eine einfache HTTP-Schnittstelle zur Formulierung von beliebig komplexen Anfragen, in welchen zeitliche Parameter und die Art der Aggregation spezifiziert werden können. In der ersten Realisierung haben wir uns auf einfache statistische Auswertungen, wie den durchschnittlichen Zeitaufwand aller Lernenden eines Kurses in einem bestimmten Zeitraum, beschränkt. Komplexere Auswertungen, wie die Analyse der Lernaktivitäten von Teilgruppen (z.B. der Gruppe „ähnlicher“ Lernender), sind Gegenstand unserer weiterführenden Arbeiten. Im unteren Bereich der Semesteransicht befindet sich zudem eine Visualisierung der Antworten zu einem der Items des oben beschriebenen Wochenrückblick-Fragebogens.

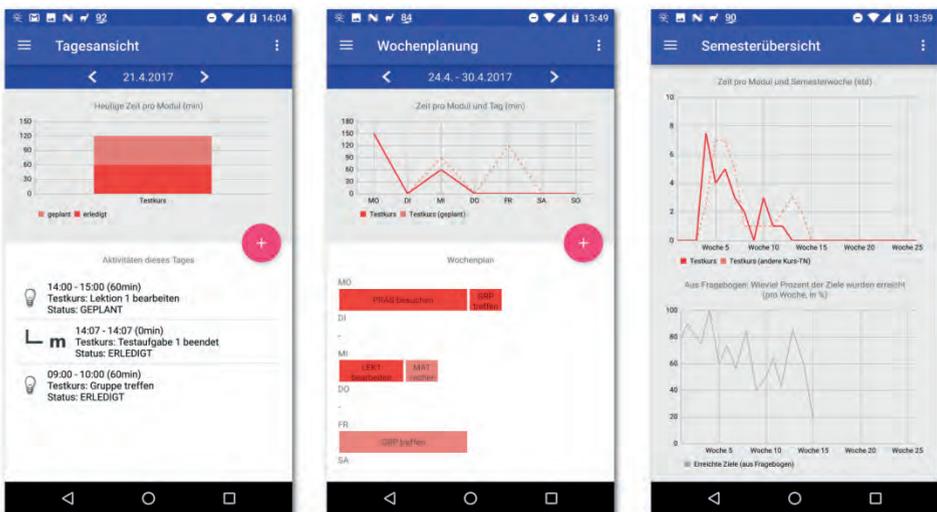


Abb. 3: Verschiedene Visualisierung von Lernaktivitäten in der Lerntagebuchanwendung

Die Lerntagebuchanwendung ist jedoch keine rein datenkonsumierende Komponente. Sie sendet ihrerseits Informationen über geplante sowie durchgeführte Lernaktivitäten an den LRS. Wir verwenden die durch xAPI vorgesehenen Erweiterungsmöglichkeiten der Activity Statements, um zu kennzeichnen, ob es sich um eine geplante oder durchgeführte Lernaktivität handelt. Die Informationen über durchgeführte Lern-

aktivitäten werden zur Berechnung der Peer-Statistiken verwendet, während die Informationen über geplante Lernaktivitäten für die Selbstregulationsforschung oder zur Gestaltung von Interventionen von Interesse sein können.

5 Zusammenfassung und Ausblick

Lernen findet heute in vielfältigen Formen und zunehmend selbstreguliert statt, was mit Vorteilen aber auch Herausforderungen für Lernende verbunden ist. Im Hochschulbereich werden neben einem zentralen LMS individuell weitere Lernanwendungen genutzt. Die Auswertung von Daten einer einzelnen Lernanwendung führt in solchen offenen Lernumgebungen zu unvollständigen Informationen über Lernende. Für personalisierte Adaptions- und Interventionsmechanismen bedarf es daher einer Integration von Daten aus verschiedenen Lernanwendungen.

In diesem Beitrag haben wir beispielhaft eine offene Lernumgebung, bestehend aus einem LMS und einer Lerntagebuchanwendung, vorgestellt und gezeigt, wie die Sammlung und Auswertung von Daten in offenen Lernumgebungen - unter Wahrung der Privatheit - realisiert werden kann. Die Lerntagebuchanwendung unterstützt Lernende beim selbstregulierten Lernen, indem sie Funktionen für die Planung und die Reflexion des eigenen Lernverhaltens zur Verfügung stellt und einen Vergleich mit Peers herstellt. Sie nutzt die zentral gespeicherten Daten um durchgeführte Lernaktivitäten teilweise automatisch zu erfassen. Gleichzeitig liefert sie auch selbst Daten, auf die durch den LRS zugegriffen werden kann.

Die in diesem Beitrag dargestellten Ergebnisse sind Ausgangspunkt für weitere Arbeiten. Basierend auf der hier nachgewiesenen Umsetzbarkeit einer zentralen Datensammlung in offenen Lernumgebungen, sollen weitere Lernanwendungen in das Szenario integriert werden. Auch Komponenten zur Datenanalyse und Visualisierung für Lehrende sollen dabei in Betracht gezogen werden. Weiterhin soll die automatische Erkennung von Lernaktivitäten durch das Lerntagebuch weiter fortgeführt werden. Dazu haben wir in Vorarbeiten erste erfolgsversprechende Ansätze zur Klassifikation von Aktivitäten mittels Sensordaten gemacht. Ziel ist es, durch die automatische Erkennung der Aktivitäten, den Pflegeaufwand für den Lernenden zu reduzieren und die Korrektheit der Daten zu steigern. Die Gestaltung und Evaluation der Lerntagebuchanwendung sind weitere Schwerpunkte unserer zukünftigen Arbeiten. Dazu gehört auch die Gestaltung von Interventionen, z.B. wenn geplantes und tatsächliches Lernverhalten voneinander abweichen. Das beschriebene Szenario und die Lerntagebuchanwendung wurden im aktuellen Semester in einer Lehrveranstaltung eingesetzt, um erste Erkenntnisse über deren Praxistauglichkeit zu gewinnen.

Literaturverzeichnis

- [Be10] Benz, B. F.: Improving the Quality of E-Learning by Enhancing Self-Regulated Learning. A Synthesis of Research on Self-Regulated Learning and an Implementation of a Scaffolding Concept. Dissertation, TU Darmstadt, 2010.
- [Bo99] Boekaerts, M.: Self-regulated learning: where we are today. *International Journal of Educational Research* 31/6, S. 445-457, 1999.
- [DRT14] Diaconita, I.; Rensing, C.; Tittel, S.: Getting the Information You Need, When You Need It: A Context-aware Q&A System for Collaborative Learning. In: *European Conference on Technology Enhanced Learning*, Springer, S. 410-415, 2014.
- [Dy13] Dyckhoff, A. L. et al.: Supporting Action Research with Learning Analytics. In: *Proc. of the 3rd Int. Conf. on Learning Analytics and Knowledge*, ACM, S. 220-229, 2013.
- [GH07] Gläser-Zikuda, M.; Hascher, T.: Lernprozesse dokumentieren, reflektieren und beurteilen. *Lerntagebuch und Portfolio in Bildungsforschung und Bildungspraxis*. Klinkhardt, 2007.
- [Ki04] Kirchhöfer, D.: *Lernkultur Kompetenzentwicklung: Begriffliche Grundlagen*. Arbeitsgemeinschaft Betriebliche Weiterbildungsforschung, 2004.
- [Ko16] Konert, J. et al.: PeerLA - Assistant for individual learning goals and self-regulation competency improvement in online learning scenarios. In: *Proc. of the 16th Int. Conf. on Advanced Learning Technologies*, IEEE, S. 52-56, 2016.
- [KR16] Kevan, J. M.; Ryan, P. R.: Experience API: Flexible, Decentralized and Activity-Centric Data Collection. *Technology, Knowledge and Learning* 21/1, S. 143-149, 2016.
- [RN02] Rambow, R.; Nückles, M.: Der Einsatz des Lerntagebuchs in der Hochschullehre. *Das Hochschulwesen* 50/3, S. 113-120, 2002.
- [Ro13] Romero, C. et al.: Web Usage Mining for Predicting Final Marks of Students That Use Moodle Courses. *Computer Applications in Engineering Education* 21/1, S. 135-146, 2013.
- [Rö17] Röpke, R.: A Pseudonymity based Identity Management to support Learning Analytics in Open Learning Environments. Masterarbeit, TU Darmstadt, 2017.
- [Sc09] Scholl, P. et al.: Implementation and Evaluation of a Tool for Setting Goals in Self-Regulated Learning with Web Resources. In: *European Conference on Technology Enhanced Learning*, Springer, S. 521-534, 2009.
- [SM11] Schulmeister, R.; Metzger, C., Hrsg.: *Die Workload im Bachelor: Zeitbudget und Studierverhalten: Eine empirische Studie*. Waxmann Verlag, 2011.
- [SW06] Schmitz, B.; Wiese, B. S.: New perspectives for the evaluation of training sessions in self-regulated learning: Time-series analyses of diary data. *Contemporary Educational Psychology* 31/1, S. 64-96, 2006.
- [Wi14] Wise, A. F.: Designing Pedagogical Interventions to Support Student Use of Learning Analytics. In: *Proc. of the 4th Int. Conf. on Learning Analytics and Knowledge*, ACM, S. 203-211, 2014.

Geführte Auswahl von IT-basierten Lehr/Lern-Werkzeugen

Tommy Kubica¹, Tenshi Hara¹, Iris Braun¹, Felix Kapp², Alexander Schill¹

Abstract: Universitäre Lehrveranstaltungen (LV) in den MINT Fächern profitieren von der Einführung von IT-basierten Lehr/Lern-Werkzeugen (ILL), beispielsweise Classroom Response Systemen (CRS). In Kombination mit selbstreguliertem Lernen bleibt das Lernen nicht auf das Wissen *über* MINT beschränkt, sondern wird auf das Lernen *durch* MINT erweitert. Nachteilig hat sich die Einführung von ILL auf den Vorbereitungsaufwand der Lehrenden ausgewirkt. Nicht nur muss der funktionale Umfang der ILL den Lehrenden bewusst sein, idealerweise verstehen sie auch alle zugrundeliegenden lernpsychologischen und didaktischen Konzepte. Derzeit existieren nur wenige Lehr/Lern-Plattformen (LLP), die die Lehrenden bei der Auswahl von ILL unterstützen, beispielsweise einschätzen, ob ein CRS-Einsatz geeignet und sinnvoll ist. Keines der existierenden Auswahlverfahren bietet dabei umfangreiche Erklärungen der didaktischen und lernpsychologischen Hintergründe oder ihrer Kompatibilität. Deshalb sind Lehrende häufig mit der ILL-Auswahl überfordert. Diese Probleme adressierend wurde ein Prototyp mit geführter und erklärender ILL-Auswahl implementiert. Alle Auswahlvorschläge werden bereits zum Zeitpunkt der Erstellung von LV durch einen simplen Algorithmus bewertet und führen zu individuellen Bewertungen für einzelne Lehrende. Gleichzeitig werden den Lehrenden bei Bedarf einblendbare, umfangreiche Hinweise und Erklärungen angeboten.

Keywords: Geführte Auswahl, Lehr/Lern-Werkzeuge, Lehr/Lern-Plattformen, Auditorium Mobile Classroom Service

1 Problemstellung

Auf Grund der zunehmenden Digitalisierung sollten Hochschulen Lehrveranstaltungen (LV) nicht mehr als klassische Vorlesungen, Übungen, Seminare, Tutorien, etc. anbieten. Stattdessen sollten IT-basierte Lehr/Lern-Werkzeuge (ILL) (mit)verwendet werden, beispielsweise Audience Response Systeme (ARS) [Ma09, We13], Echtzeitevaluationssysteme, oder ganz generell Evaluationen an sich [Ki07]. Diese werden oft unter der Bezeichnung *Classroom Response System (CRS)* zusammengefasst. Im Bereich der MINT-Fächer (Mathematik, Informatik, Naturwissenschaft und Technik) ist eine der Kernideen dabei, MINT-Studierenden nicht nur Wissen *über* sondern auch *durch* MINT zu vermitteln³. Die angesprochenen ILL sind dabei nur ein Anfang.

Herausfordernd stellt sich in der Regel bereits die Auswahl geeigneter ILL dar. Häufig

¹ Technische Universität Dresden, Fakultät Informatik, Institut für Systemarchitektur, Professur für Rechnernetze, 01062 Dresden, {vorname.nachname}@tu-dresden.de

² Technische Universität Dresden, Fakultät Mathematik und Naturwissenschaften, Fachrichtung Psychologie, Professur für Psychologie des Lehrens und Lernens, 01062 Dresden, felix.kapp@tu-dresden.de

³ Natürlich können diese Überlegungen auch auf Fächer außerhalb von MINT angewendet werden, was die entsprechenden LV entsprechend modernisiert und ansprechende macht.

sind Lehrende vom Funktionsumfang, der Kompatibilität von ILL untereinander, oder den zugrundeliegenden didaktischen und lernpsychologischen Konzepten überfordert. Dies führt oftmals dazu, dass Lehrende eigentlich besonders gut für ihren Bedarf geeignete ILL falsch oder gar nicht verwenden. Zusätzlich sind Lehrende manchmal ignorant gegenüber besonders geeigneter oder besonders ungünstiger ILL-Kombinationen. Das in Kombination mit der bereits angesprochenen funktionalen Überforderung führt zu einer defensiven Haltung gegenüber dem Einsatz von ILL, was natürlich dem eigentlichen Ziel, dem vermehrten ILL-Einsatz, abträglich ist. Auch steigt die Arbeitsbelastung der Lehrenden, was unnötig und vermeidbar ist. Diese Situation wird dadurch verschärft, dass Auswahlüberlegungen nicht nur einzelne LV berücksichtigen sollten, sondern sich über ganze LV-Reihen erstrecken sollten. Schließlich ändert sich der Einsatzkontext von ILL über den Verlauf eines Semesters.

2 Verwandte Arbeiten

Das Problem des vermeidbaren Arbeitsmehraufwandes und der Überforderung der Lehrenden ist weder unbekannt noch neu. Tatsächlich handelt es sich hierbei sogar eher um ein allgemeines Problem der Software-Entwicklung. Alle Best-Practices sagen, dass im System neue Nutzer in die Nutzung eingeführt werden müssen, sei es durch Reduktion des initialen Funktionsumfangs, durch geeignete visuelle Filterung (Hervorheben/Ausblenden von Funktionen), oder durch ein gezieltes On-Board-Tutorial. In Bezug auf ILL kann dieses Wissen aus der Software-Entwicklung natürlich angewendet werden: Hauptsächlich wird auf die Reduktion auf eine *fixierte Funktionsvorauswahl*, nicht näher erklärte *Funktionsauswahl*, *erklärte Funktionsauswahl* und *geführte Funktionsauswahl* (oder *Vorschlagsbasierte Funktionsauswahl*) gesetzt. Letzteres wurde in *Auditorium Mobile Classroom Service (AMCS)* [Ka14, Ha15, KBH16] umgesetzt.

Existierende Lehr/Lern-Plattformen (LLP) den eingeführten Auswahltypen zuzuordnen ist sinnvoll, denn es sollten nicht in allen Lehrformen immer alle ILL bereitgestellt werden [LVC14]. Somit ergibt sich als natürlicher Diskriminator der Funktionsumfang.

2.1 Fixierte Funktionsauswahl

Viele existierende LLP-Systeme sind auf ganz eingeschränkte Szenarien festgelegt. Ihre Funktionalität ist daher auf eine Vorauswahl ausgewählter ILL, die für das Szenario relevant sind, beschränkt. *Backchannel*⁴ beispielsweise unterstützt nur Instant Feedback, wohingegen *polleverywhere*⁵ zusätzlich das Beantworten von Fragen ermöglicht.

Insbesondere *polleverywhere* zeigt, dass es sinnvoll ist, den Funktionsumfang dem Szenario entsprechend zu reduzieren. Das Mantra lautet „Qualität vor Quantität“. Im

⁴ <https://backchannel.cnc.io/> – abgerufen am 1. März 2017

⁵ <https://polleverywhere.com/> – abgerufen am 1. März 2017

Vergleich zu anderen Q&A-Systemen unterstützt *polleverywhere* auch unterschiedliche Fragentypen (bspw. Brainstorming) und unterschiedliche Teilnahmewege (Webseite, SMS, Twitter). Echtzeitergebnisse können auf der Webseite geteilt werden, oder direkt in *Powerpoint*, *Keynote* oder *Google Slides* eingebettet werden. Der Fragentyp und Einbettungsgrad kann durch die Lehrenden festgelegt werden, oder manuell pausiert, fortgesetzt oder beendet werden.

Da Systeme mit fixierter Funktionsvorauswahl auf ihr Szenario zugeschnitten sind, treten in der Regel keine Inter-ILL-Kompatibilitätsprobleme auf, solange alle LV einer LV-Reihe demselben Szenario entsprechen.

2.2 Funktionsauswahl

Systeme mit über die erwähnten Echtzeitrückmelde- und Q&A-Funktionen hinausgehendem Funktionsumfang erlauben in der Regel die Auswahl besagter Funktionen durch die Lehrenden.

*Letsfeedback*⁶ erlaubt das Erstellen von *Hörsaalfragen* oder *Willkommensnachrichten* (siehe Abbildung 2a). Ähnlich wie bei *polleverywhere* können die Hörsaalfragen durch pausieren/fortsetzen der Funktion kontrolliert werden. Auf Grund des sehr begrenzten Funktionsumfangs und die selbsterklärenden Namen kann die Auswahl der Funktionen als *intuitiv* angesehen werden.

Bei *SMILE*⁷ findet die ILL-Auswahl ähnlich statt. Das Funktionsauswahlmenü erlaubt es, sogenannte *Module* zu aktivieren oder deaktivieren. Jedes Modul enthält unterschiedliche funktionale Aspekte von ILL. Nach der Aktivierung eines Moduls wird ein neuer Tab eingeblendet, innerhalb dessen das Modul konfiguriert werden kann. Dennoch ist der Funktionsumfang im Allgemeinen eher begrenzt, und mehrdeutige Namensgebung (mangels Namenskonvention) führt zur Verwirrung des nicht eingearbeiteten Nutzers.

Systeme mit nicht näher erklärter Funktionsauswahl sind auf die zügige Erstellung einzelner LV ausgelegt. Die Auswirkungen auf LV-Reihen werden nicht berücksichtigt. Das macht die Systeme anfällig für ILL-Kompatibilitätsprobleme, die sich im Verlaufe eines Semesters ergeben können. Nach unserem besten Wissen berücksichtigt keines der Systeme Konzepte für LV-Reihen oder warnt vor ILL-Kompatibilitätsproblemen.

2.3 Erklärte Funktionsauswahl

Ein *SMILE* sehr ähnliches System ist *Tweedback*⁸, wobei es alle ILL im Funktionsauswahlmenü beschreibt (Abbildung 1). Die Beschreibungen sind auf funktionale Aspekte

⁶ <https://letsfeedback.com/en/home/> – abgerufen am 1. März 2017

⁷ <https://www.smile.informatik.uni-freiburg.de/> – abgerufen am 1. März 2017

⁸ <https://www.tweedback.de/> – abgerufen am 1. März 2017

beschränkt; didaktische oder lernpsychologische Hintergründe fehlen. In Bezug auf die verfügbare Funktionalität ist diese Form der Beschreibung aber gut gelungen.

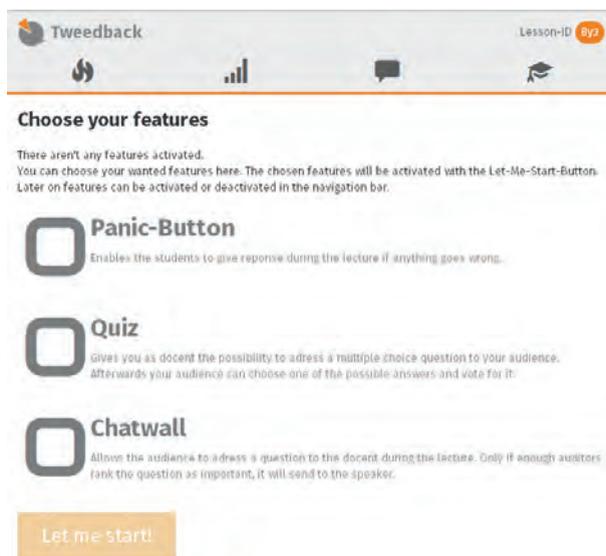
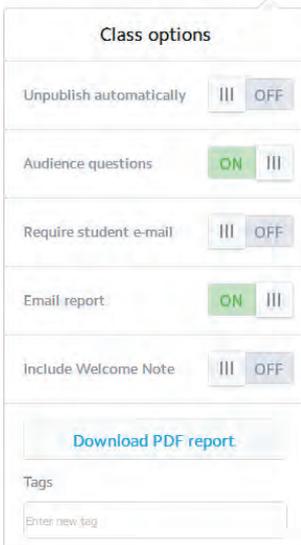


Abb. 1: Beschreibung der Abbildung

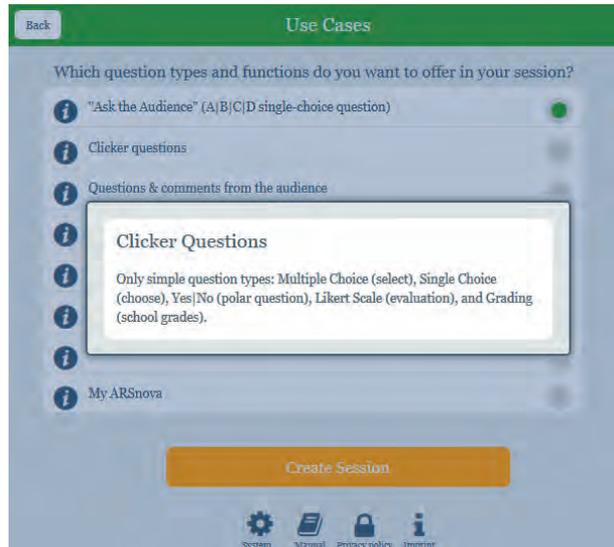
Ein funktional komplexes und umfangreiches System ist *ARSnova*⁹, welches die ILL-Auswahl anhand verschiedener Anwendungsfälle umsetzt (Abbildung 2b). Jeder Anwendungsfall hat einen mehr oder weniger intuitiven Bezeichner und präsentiert weitere Informationen als Modal-Nachricht nachdem ein Hilfe-Symbol angeklickt wurde. Diese Beschreibungen enthalten teilweise auch einen didaktischen Kontext für den jeweiligen Anwendungsfall.

Neben den vorgegebenen Anwendungsfällen erlaubt *ARSnova* das Erstellen eigener individueller Funktionsgruppen. Solche Gruppen können durch Auswahl von ILL aus der Liste aller Funktionen und Erweiterungen erstellt werden. Bei den individuellen Funktionsgruppen werden jedoch keinerlei funktionale oder didaktische Hinweise gegeben, was ungünstig ist, da den Lehrenden durchaus abstrakte und unbekannte Konzepte vorgesetzt werden können. Ihr funktionaler Umfang und ihre Ziele können auf Basis der Funktionsname keineswegs intuitiv hergeleitet werden. Auf diese Art können LV-Reihen-überspannende Effekte nicht berücksichtigt werden.

⁹ <https://arsnova.eu/mobile/> – abgerufen am 1. März 2017



(a) Auswahlmenü



(b) Anwendungsfallauswahl und -beschreibung

Abb. 2: Letsfeedback (a) und ARSnova (b)

Während *Tweedback* lediglich die genannten Beschreibungen gegenüber *Letsfeedback* ergänzt, bietet ARSnova eine Fülle zusätzlicher Funktionen. Beispielsweise ist die *interaktive Präsentation* ein probates Mittel, um LV als Peer Instruction [Ma97] umzusetzen. Negativ daran ist jedoch die exklusive Beschränkung auf Unterfunktionen beziehungsweise Teilaspekte von ILL. Während Publikumsfragen ähnlich wie bei „Wer wird Millionär“ in bestimmten Situationen sinnvoll sein können, möchten Lehrende in anderen Situation vielleicht lieber Evaluationsfragen stellen. Nichtsdestotrotz gibt es in *ARSnova* keinen Standardanwendungsfall in dem beide Funktionen verfügbar sind. Beim Erstellen einer entsprechenden individuellen Funktionsgruppe müssen Lehrende mangels Beschreibung der ILL erst erraten, dass Hörsaalfragen zur Umsetzung von Publikumsfragen und Evaluationsfragen dienen und doppelt ausgewählt werden müssen.

2.4 Geführte Funktionsauswahl

Wie in den vorangegangenen Abschnitten ersichtlich geworden ist, bieten viele Systeme eine Funktionsauswahl für individuelle LV. Es werden teilweise auch Beschreibungen und Hinweise zu den ILL oder ihren Anwendungsfällen angeboten. Keines der betrachteten Systeme berücksichtigt jedoch die Beschränkungen der ILL oder versucht sich an einer sinnvollen Vorauswahl wie beispielsweise in [LVC14] gefordert. Im Allgemeinen werden Lehrende im Stich gelassen sobald es darum geht, einzuschätzen wie sinnvoll bestimmte Funktionen in ihren LV eingesetzt werden können. Betrachtungen, die mehrere LV oder ganze LV-Reihen überspannen gibt es gar keine.

Nach unserem besten Wissen existiert derzeit kein weiteres System neben unseren Prototypen, das Lehrende bei der ILL-Auswahl in der Art unterstützt, dass sinnvolle Einschätzungen sowohl bezüglich einzelner LV als auch ganzer LV-Reihen unter Berücksichtigung des didaktischen und lernpsychologischen Kontexts und von ILL-Kompatibilität gibt. Insbesondere im Vergleich zu *ARSnova* unterstützt unser Prototyp Lehrende immer auch unter Berücksichtigung von Kompatibilität und dem Zeitpunkt im Semester (insbesondere wann innerhalb einer LV-Reihe eine LV stattfindet).

3 Prototyp

Um eine Plattform für unsere Forschungsarbeiten im Gebiet der technologieunterstützten Lehre zu haben, die aktuelle Entwicklungen der Lernpsychologie, Didaktik und Informatik umsetzen kann, wurde in den vergangenen Jahren intensiv am bereits genannten *Auditorium Mobile Classroom Service*¹⁰ (*AMCS*) entwickelt. Die Hauptfunktionen von *AMCS* wurden bereits intensiv untersucht und kontinuierlich weiter entwickelt [Ka14, Ha15, KBH16]. *AMCS* soll dabei eine sinnvolle Ergänzung sowohl für Lehrende als auch die Lernenden sein. Aktive, konstruktive und hochgradig individuelle Lernprozesse nach [Se03] sollen dabei umgesetzt werden. Mit einem der nächsten Updates wird *AMCS* auch aktuelle Ergebnisse in Bezug auf Tutorien berücksichtigen. *AMCS* konzentriert sich konzeptuell insbesondere auf didaktische Konzepte wie Peer Instruction [Ma97] und Audience Response Systeme [Ma09, We13]. Nichtsdestotrotz ist *AMCS* in einer lernpsychologischen Basis wohlfundiert, welche den individuellen Lernprozess der Studierenden in LV(-Reihen) beschreibt. Die umgesetzten Funktionen leiten sich von Modellen des selbstregulierten Lernens ab (beispielsweise [Ha01] oder [Zi00]). Lehrende können Lernaufgaben erstellen, Umfragen durchführen und Nachrichten an individuelle Studierende automatisiert schicken lassen. Solche Interventionen werden während LV nach definierten Regeln versendet. Lehrende werden dadurch zu Konstrukteuren interaktiver, technologieunterstützter Lernumgebungen. Des Weiteren wird die Durchführung umfassender Evaluationen unterstützt [KBH16]. Leider ist auch *AMCS* ein für neue Lehrende zu komplexes und zu umfangreiches System geworden, welches nicht mehr intuitiv zu verstehen war (vergleiche dazu *ILIAS*¹¹ und *ARSnova*). Aus diesem Grund wurde im Rahmen einer Master-Arbeit eine geführte Funktionsauswahl ergänzt.

AMCS basiert auf einem zentralen Web-Server (Ruby on Rails) mit einer Datenbank für die Informationspersistierung. Die Darstellung von Informationen und die komplette Nutzerinteraktion finden in Web-Browsern auf Computern oder in Apps auf Smartphones (Android und iOS) statt. Der primäre Nutzungszugang ist das eigene Gerät der Nutzer (BYOD-Ansatz – Bring your own Device), kombiniert mit dem Second-Screen-Ansatz, bei dem das Gerät einen parallelen (meist optionalen) Kanal zur eigentlichen LV darstellt. Die Nutzung selbst ist anonym unter Angabe eines selbst

¹⁰ <https://amcs.website/>

¹¹ <https://ilias.uni-giessen.de/ilias/> – abgerufen am 1. März 2017

ausgewählten Pseudonyms und eines selbst ausgewählten Passwortes möglich; Nutzer können sich jederzeit unter Angabe abweichender Angaben eine neue „Identität“ geben. Dadurch ist insbesondere bei der Beantwortung von Lernfragen, der Teilnahme an Evaluationen und beim Stellen eigener Fragen die Anonymität garantiert. Alle Informationen werden verschlüsselt gespeichert, insbesondere die Kennwörter nur als Hash per bcrypt mit Einmal-Salt.

Aktuell umfasst *AMCS* sieben aktiv verwendete ILL, nämlich Aggregation von Interessen und persönlichen Zielen, Lernaufgaben, meta-kognitive Prompts, kognitive Prompts, Bereitstellung zusätzlichen Materials, Scripted Discussions und LV-Evaluationen [Ka14, Ha15, KBH16]. Zusätzlich werden derzeit zwei weitere ILL getestet: grafisches Feedback und grafische Diskussionen [Ha17]. Alle integrierten ILL zielen darauf ab, Studierende beim individuellen Lernprozess zu unterstützen. Nach [Zi00] werden Studierende mit Herausforderungen in allen Phasen ihres Lernens konfrontiert, weshalb geeignete Unterstützung den Lernprozess erfolgreicher gestalten kann. Studierende unterscheiden sich bezüglich ihrer individuellen Ausgangssituation, ihrer Ziele, ihrer Wahrnehmung, der Wissensverarbeitung, etc. Ihre Lernaktivitäten evaluieren zu können und dann individuell angepasste Unterstützung bieten zu können ist ein massiver Vorteil, den *AMCS* umzusetzen versucht.

4 Geführte Auswahl – Nutzungsoberfläche

Die neu eingeführte Nutzeroberfläche für die geführte Auswahl (GS-GUI) hat als zentrales Merkmal eine Reihe einfach zu beantwortender Fragen. Ähnlich einem Myers-Briggs-Test wird durch möglichst wenige gezielte Fragen sehr schnell die Menge geeigneter ILL determiniert, indem explizit auf die Randbedingungen der LV(-Reihen) eingegangen wird (Abbildung 3). Während die Lehrenden die Fragen beantworten werden ihnen Erklärungen und Einschätzungen angezeigt, welche die Funktionalität und didaktischen Konzepte der ILL darlegen. Zusätzlich zeigt das GS-GUI weitere Informationen an sobald ein ILL manuell oder automatisch ausgewählt wird. Diese Anzeigen beinhalten insbesondere Informationen über die Kombinierbarkeit von ILL; wie sie sich gegenseitig befördern oder behindern können (in der Regel auch anhand eines Beispiels). Am Ende wird den Lehrenden eine Einschätzung zur Eignung einer ILL-Auswahl für ihre individuelle LV angezeigt (Abbildung 4); erklärende Informationen sind weiterhin abrufbar. Derzeit ist unser Prototyp nicht in der Lage LV-Reihen anzulegen, aber ist sich des Kontextes einzelner LV innerhalb einer LV-Reihe bewusst.

Die Ergebnisse der geführten Auswahl werden durch einen simplen, erweiterbaren Algorithmus berechnet auf Basis von Pflichtfaktoren (derzeit fünf; es können weitere definiert werden), nämlich der Publikumsgröße, Verfügbarkeit von Hörsaaltechnik (Beamer, etc.), verfügbare Vorbereitungszeit der Lehrenden, verfügbare Vor-Ort-Zeit für ILL und dem Zeitpunkt im Semester. Jeder der Pflichtfaktoren wird auf eine von drei Kategorien abgebildet (feiner granulare Einteilung ist technisch möglich, wird aber erst

später ergänzt). Beispielsweise bildet die Publikumsgröße derzeit ab auf: „1 bis 25 Studierende“ (typisches Tutorium), „25 bis 50 Studierende“ (typisches Seminar oder kleine Vorlesung) und „mehr als 50 Studierende“ (allgemeine Vorlesung). Die Kategorien können auf die Randbedingungen unterschiedlicher Bildungseinrichtungen angepasst werden.

Abb. 3: Abfrage von LV-Parametern

Zusätzlich zu den Pflichtfaktoren können drei optionale Parameter abgefragt werden (wiederum können mehr später definiert werden). Diese optionalen Parameter decken die Einstellungen der Studierenden gegenüber Teilnahme und der Lehrenden gegenüber studentischer Involvierung, und den von den Lehrenden erwartete Anteil der ILL nutzenden Studierenden ab. Auch hier teilt der Algorithmus in drei Kategorien ein.

Nach der initialen Kategorisierung und Berechnung einer ersten Auswahl verfeinert der Algorithmus das Ergebnis, indem die Erfahrungen der Lehrenden und Studierenden mit der LLP (wie oft sie das System bereits verwendet haben) und die Beziehungen zwischen den einzelnen in der Auswahl vorhandenen ILL (ob sie sich gegenseitig fördern oder behindern) berücksichtigt werden. Diese Beziehungen sind wiederum in drei Kategorien eingeteilt: solche, die sich gegenseitig ausschließen; solche, die oft gemeinsam verwendet werden (schwache Korrelation); solche, die sich gegenseitig bedingen (starke Korrelation). Erst nach dieser letzten Reduktion der Auswahl wird das Ergebnis des Algorithmus' für die Ausgabe in der GS-GUI aufbereitet (Abbildung 4).

Am Ende haben die Lehrenden die Möglichkeit, bei Bedarf die Auswahl individuell anzupassen, um sie ihren eigenen Vorlieben Rechnung zu tragen. Beispielsweise können Lehrende, die Diskussions-basierte ILL nicht mögen selbige abwählen. Jede Änderung der vorgeschlagenen Auswahl wird durch die GS-GUI durch eine Benachrichtigung mit

Kommentaren, Erklärungen und Empfehlungen bestätigt. Im Allgemeinen sind diese Benachrichtigungen stark auf didaktische Kontexte fokussiert, aber es werden auch funktionale Aspekte der ILL aufgeführt, was wichtig bei ILL mit starker Korrelation ist, da die Abwahl einer Funktion automatisch zur Abwahl weiterer ILL führen kann.

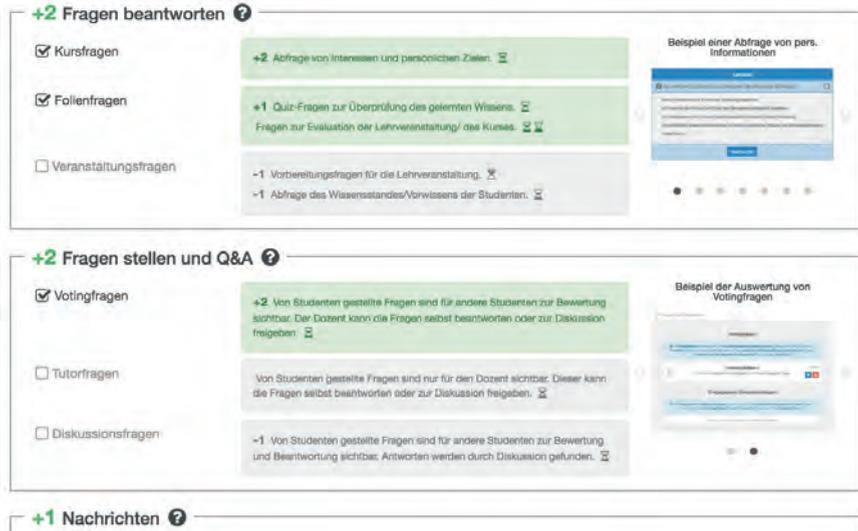


Abb. 4: Auszug der ILL-Auswahl basierend auf den Angaben der Lehrenden

5 Experimente

5.1 Experimentelles Setup

Unsere Untersuchungen wurden in Bezug auf Funktionalität, Nutzbarkeit und Zuverlässigkeit der Auswahl durchgeführt. Die Erfassung erfolgte auf Basis eines System Usability Scales (SUS) [NM90], Thinking-Aloud-Sitzungen und der tatsächlichen Ergebnisausgabe. Alle Tests mit Nutzerinteraktionen wurden auf der gleichen physischen Maschine durchgeführt, einem 13.3" Apple MacBook Pro mit einer 3.1 GHz Core i7 CPU und 16 GB RAM. Die Tests wurden unter OS X 10.11.3 in einer 64bit-Instanz von Google Chrome 52.0.2743.116 bei einer Auflösung von 1440x900 px durchgeführt. 11 Probanden (davon 5 „Experten“, die das System bereits kannten) haben in zwei Iterationen an den Tests teilgenommen. Jeder Test wurde gegen Mittag durchgeführt und dauerte etwa eine Stunde. Die Systemausgaben, insbesondere die Ergebnisse des Algorithmus, wurden notiert, bevor das System nach jedem Test auf den Ausgangszustand zurückgesetzt wurde. Die Probanden nutzten das System dabei in drei fiktiven Szenarien einer gut besuchten Mathematik-LV-Reihe: die Einführungsvorlesung, ein

kleineres Tutorium mitten im Semester, sowie das Abschlusssseminar kurz vor der Prüfung am Semesterende. Dabei wurde beim Durchspielen per Thinking-Aloud das Vorgehen der Probanden und im Anschluss durch gezielte Befragung bezüglich der Erwartungserfüllung durch die Systemreaktionen eine Einschätzung der Qualität der Funktionsauswahl erfasst. Abschließend füllte jeder Proband einen Fragebogen zur Ermittlung des SUS aus.

5.2 Ergebnisse

Auf Grundlage der erhaltenen Rückmeldungen konnten Best-Practices erfasst werden. Die wichtigste Information für Lehrende ist demnach, wie viel Zeit der Einsatz eines ILL von ihnen abverlangt, sowohl in der Vorbereitung, als auch vor-Ort beim eigentlichen Einsatz. Daher muss ganz klar erfasst und kommuniziert werden, wie viel Zeit Lehrende willig sind in die Vorbereitung zu investieren, wie viel Zeit vor-Ort bereitsteht und wie viel davon sie bereit sind „zu opfern“. Unabhängig davon wie selbsterklärend ein ILL zu sein scheint, Erklärungen und Hinweise müssen *immer* verfügbar sein. Diese sollen auch Einschätzungen der Auswirkungen auf den Auswahlalgorithmus enthalten. Für den Fall, dass ein ILL mit einem anderen kollidiert, muss ein klarer, von der Nutzung abratender Hinweis erfolgen, idealerweise als modale Nachricht. In der Umkehrung sollen gut gemeinsam funktionierende ILL besonders hervorgehoben werden. Basierend auf den Parametern einer LV (Zeitpunkt im Semester, Raumgröße, etc.) muss ein visuell klarer Hinweis erfolgen, wie sich diese positiv oder negativ auf den Auswahlalgorithmus auswirken. Dies entspricht einer Aussage über die didaktische Sinnhaftigkeit, ohne den Lehrenden die didaktischen Hintergründe im Detail zu erklären. – Der Auswahlprozess muss sich so transparent wie möglich präsentieren. Er darf keine „Black Box“ sein.

Zwischen den beiden Iterationen wurde der LLP-Prototyp entsprechend angepasst. Jedoch stieg der SUS wider Erwarten nicht, sondern sank von 88.3% in der ersten auf 80.6% in der zweiten Iteration. Wegen insignifikanter Korrelationstests, muss dies in Zukunft näher untersucht werden, aber für den Augenblick ist die Annahme, dass die Ursache in der Zusammensetzung der Probandengruppen liegt. Die erste Iteration deckte nur Experten ab; die zweite hatte einen signifikanten Anteil dem System neuer Nutzer.

Aus algorithmischer Sicht waren die Ausgaben des Systems erwartungskonform. Für alle drei Szenarien waren die Auswahlen in der Einschätzung der Probanden korrekt.

Im ersten Szenario wurden alle Hörsaalfragen-basierten ILL mit Ausnahme der Vorbereitungsfragen in die Auswahl aufgenommen. Dies ist korrekt, da die Annahme ist, dass die Studierenden kein Vorwissen haben und auch nicht wissen, dass es die LLP mit den Fragen gibt (sie können sich also nicht im System vorbereiten). Die Wiederholung von Stoff aus vorangegangenen LV ist auch nicht möglich, da es keine solchen gibt. Der Fortschritt im Lernprozess kann ebenso wenig bestimmt werden. Auf Grund des zu erwartenden großen Publikums hat der Algorithmus erwartungskonform von der Nutzung von Tutorfragen (Studierende können Fragen direkt an die Lehrenden schicken) oder Diskussionen abgeraten. Das Spam-Risiko und Ablenkung wäre einfach zu groß.

Im zweiten Szenario wurde von der Nutzung jeglicher interessenbasierten sowie von lernfortschritt-basierten ILL abgeraten. Dies ist sinnvoll da Tutorien den Zweck der Wissensverfestigung erfüllen, worin das in Vorlesungen vermittelte Wissen angewendet und vertieft werden soll. Zusätzlich wurde die Verwendung von CRS-Funktionen nicht empfohlen, da diese bei kleinen Gruppen wenig(er) Sinn machen.

Im dritten Szenario wurden im Ergebnis rückmelde- und fragenbasierte ILL besonders empfohlen. Dies ist erwartungskonform da Studierende in der Prüfungsvorbereitung durchaus viele Fragen haben können. Dies umfasst sowohl Fragen zu individuellen Themen, die die Studierenden direkt mit den Tutoren klären wollen, aber auch Diskussionen, um mit Kommilitonen einen regen Austausch betreiben zu können. Stark abgeraten wurde von Evaluationsfunktionen, da zu diesem späten Zeitpunkt kein Nutzen für evaluatives Feedback besteht. Zukünftige Iterationen der LV könnten zwar verbessert werden, aber den aktuellen Studierenden würde sich kein Mehrwert erschließen.

6 Zusammenfassung und Ausblick

Die geführte Auswahl kann die zeitraubende Vorbereitungsphase verkürzen und verständlicher gestalten. Dem Algorithmus können didaktisch nicht versierte Lehrkräfte im Großen und Ganzen vertrauen. Nichtsdestotrotz zeigt eine nähere didaktische Analyse, dass die Ergebnisse durchaus kontrolliert werden sollten. Die geführte Auswahl ist dennoch eine sinnvolle Ergänzung für Lehrende, da schnell eine Basisauswahl getroffen werden kann, welche anschließend lediglich verfeinert wird. Dies verkürzt die benötigte Vorbereitungszeit. Der im allgemeinen positive SUS (über 80%) zeigt, dass der gewählte Ansatz durchaus eine geeignete Hilfestellung für die Lehrenden ist.

In Zukunft wird ein interaktives Tutorial benötigt. Obwohl der Prototyp auf eine Führung der Nutzer und eine Unterstützung bei der Auswahl von ILL ausgelegt ist, sind die ILL an sich immer noch zu komplex. Es ist jedenfalls nachteilig, wenn Lehrende bei der Auswahl der ILL unterstützt werden, dann aber bei der Nutzung keine Hilfe erhalten.

Das Ergänzen von Gewichten im Auswahlalgorithmus könnte die wahrgenommene Ergebnisqualität verbessern. Beispielsweise könnten die Peer Instruction bevorzugende Lehrende ein starkes Interesse daran haben, den Algorithmus zu Gunsten von Fragenbasierten ILL im Rahmen der ConcepTests zu gewichten.

Den Prototypen um die Möglichkeit eines Lehrenden-Feedbacks zu ergänzen könnte sich ebenfalls lohnen. Sobald Lehrende im System Feedback zur Sinnhaftigkeit der ILL-Auswahlergebnisse und der eigentlichen ILL-Nutzung geben können, kann der Auswahlalgorithmus in der Art abgeändert werden, dass er aus dem Feedback heraus zukünftige Ergebnisse verbessert.

Hinweis: Dieser Beitrag fasst die Ergebnisse von Tommy Kubicas Master-Arbeit zusammen und ergänzt diese um neue Erkenntnisse und notwendige Korrekturen.

Literaturverzeichnis

- [Ha01] Hadwin, Allyson Fiona; Winne, Philip H.; Stockley, Denise B.; Nesbit, John C.; Woszczyna, Carolyn: Context moderates students' self-reports about how they study. *Journal of educational psychology*, 93/3, S. 477, 2001.
- [Ha15] Hara, Tenshi; Kapp, Felix; Braun, Iris; Schill, Alexander: Comparing Tool-supported Lecture Readings and Exercise Tutorials in Classic University Settings. In: *Proceedings of the 7th International Conference on Computer Supported Education (CSEdu 2015)*, S. 244–252, 2015.
- [Ha17] Hara, Tenshi; Chen, Kaijun; Braun, Iris; Kapp, Felix: Graphicuss - Proposing a Graphical Discussion System. In: *Proceedings of the 9th International Conference on Computer Supported Education (CSEdu 2017)*, 2017.
- [Ka14] Kapp, Felix; Braun, Iris; Kördle, Hermann; Schill, Alexander: Metacognitive Support in University Lectures Provided via Mobile Devices - How to Help Students to Regulate Their Learning Process during a 90-minute Class. In: *Proceedings of the 6th International Conference on Computer Supported Education (CSEdu 2014)*, S. 194–199, 2014.
- [KBH16] Kapp, Felix; Braun, Iris; Hara, Tenshi: Evaluating Lectures Through the Use of Mobile Devices - Auditorium Mobile Classroom Service (AMCS) as a Means to Bring Evaluation to the Next Level. In: *Proceedings of the 8th International Conference on Computer Supported Education (CSEdu 2016)*, S. 251–257, 2016.
- [Ki07] Kirkpatrick, Donald: *The four levels of evaluation*, Jgg. 701, American Society for Training and Development, 2007.
- [LVC14] Leiding, Benjamin; Vetterick, Jonas; Cap, Clemens H.: *Exploring Classroom Response Systems in Practical Scenarios*. 2014, https://www.researchgate.net/profile/Benjamin_Leiding/publication/280008669_Exploring_Classroom_Response_Systems_in_Practical_Scenarios/links/55a3773a08aeafdb97bba3b.pdf, Stand: 11.07.2017.
- [Ma97] Mazur, Eric: *Peer Instruction: A User's Manual*. Prentice Hall, Upper Saddle River, NJ, Series in Educational Innovation, 1997.
- [Ma09] Mayer, Richard; Stull, Andrew; DeLeeuw, Krista; Almeroth, Kevin; Bimber, Bruce; Chun, Dorothy; Bulger, Monica; Campbell, Julie; Knight, Allan; Zhang, Hangjin: Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary educational psychology* 34/1, S. 51–57, 2009.
- [NM90] Nielsen, Jakob; Molich, Rolf: Heuristic evaluation of user interfaces. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, S. 249–256, 1990.
- [Se03] Seel, Norbert M.: *Psychologie des Lernens: Lehrbuch für Pädagogen und Psychologen*, Jgg. 8198, UTB, 2003.
- [We13] Weber, Katrin; Becker, Bernd et al.: *Formative Evaluation des mobilen Classroom-Response-Systems SMILE. E-Learning zwischen Vision und Alltag*. Waxmann, 2013.
- [Zi00] Zimmerman, B.J.; Boekarts, M.; Pintrich, P.R.; Zeidner, M.: Attaining Self-Regulation: a social cognitive perspective. *Handbook of self-regulation*, 13, S. 13–39, 2000.

Planung von Schulunterricht: Automatisches Feedback zur Reflexionsanregung über eigene Unterrichtsentwürfe

Sven Strickroth¹ & Niels Pinkwart²

Abstract: Unterrichtsplanung ist eine zugleich wichtige als auch anspruchsvolle Aufgabe und wird als Voraussetzung von gutem Unterricht gesehen. Das Erlernen der Planung ist ein Kernbestandteil der Lehrpersonenausbildung und erfordert stets Feedback von MentorInnen bzw. Dozierenden. Da nicht immer eine zeitnahe Feedbackgabe möglich ist, stellt sich die Frage, ob und wie computergeneriertes Feedback genutzt werden kann. In diesem Paper wird ein erster Ansatz für automatisches Feedback vorgestellt, der sich in einem Vergleich zu papierbasierten Leitfäden als wirkungsvoller hinsichtlich der Beachtung und Anregung zur Reflexion erwiesen hat.

Keywords: Unterrichtsplanung, Unterrichtsvorbereitung, Feedback, Reflexion

1 Einleitung

Das Planen und Vorbereiten von Unterricht ist eine zentrale Aufgabe von Lehrpersonen, denn ein effektiver, zielgerichteter Unterricht entsteht im Allgemeinen nicht von allein: Viele Faktoren, die einen Einfluss auf den Lernerfolg haben, hängen von der Lehrperson und dessen Vorbereitung ab, wie z. B. Klarheit im Hinblick auf Ziele, Inhalte, Methoden und Medien [Pe94, Zi14]. Die vermutlich am weitesten verbreitete Form, um Unterricht vorzubereiten, besteht im Erstellen von Unterrichtsskizzen bzw. -entwürfen. Dabei handelt es sich um eine essentielle, aber gleichzeitig auch kreative und anspruchsvolle Aufgabe, um Gedanken zu vorgesehenen Abläufen, Inhalten und Zielen zu ordnen. Bei der Erstellung müssen viele Aspekte zur gleichen Zeit berücksichtigt und sinnvoll zusammengestellt werden. Dazu gehören u. a. die (Fach-)Inhalte, Rahmenlehrpläne, Impulse, antizipierte Schülerantworten und die Lerngruppe mit ihrem vorhandenen Wissen und Interessen. Unterrichtsentwürfe stellen dabei in der Regel nicht nur den vorgesehenen Ablauf dar oder erfüllen den Zweck der Dokumentation (als Begründung für Entscheidungen), sondern sollen auch direkt beim Prozess des Erstellens helfen, das Geplante zu reflektieren, Optionen abzuwägen und Selbstbewusstsein aufzubauen.

Insgesamt ist für die Planung von gutem Unterricht viel Erfahrung notwendig. Daher werden angehende Lehrpersonen bereits im Studium in didaktischen Seminaren, Übungen oder Praktika mit Unterrichtsentwürfen konfrontiert, um das Planen von Unterricht zu erlernen. Speziell vor real durchzuführendem Unterricht wird von angehenden

¹ Institut für Informatik und Computational Science, Universität Potsdam, August-Bebel-Str. 89, 14482 Potsdam, sven.strickroth@uni-potsdam.de

² Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, niels.pinkwart@hu-berlin.de

Lehrpersonen oft verlangt, die Entwürfe vorher abzugeben. Ein zentraler Punkt besteht dabei im direkten Feedback von Dozierenden bzw. erfahrenen Lehrpersonen, um bereits vor der Durchführung Hinweise auf mögliche Schwierigkeiten Verbesserungen zu erhalten. Ein generelles Problem besteht jedoch darin, dass nicht immer MentorInnen zur Verfügung stehen, um zeitnah ein adäquates individuelles Feedback zu geben.

Um diesem Problem zu begegnen, wird im Folgenden ein erster Ansatz für automatisches Feedback für Unterrichtsentwürfe vorgestellt, um angehende Lehrpersonen bei der Konzeption hochwertiger Lehre zu unterstützen. Der vorgeschlagene Ansatz erzeugt Hinweise hinsichtlich der Erfüllung von Standardanforderungen sowie Auffälligkeiten z. B. bei der zeitlichen Planung und wurde in einer Studie im Direktvergleich mit herkömmlichen papierbasierten Leitfäden explorativ evaluiert. Dabei ging es vor allem darum, wie automatisches Feedback in dieser Domäne empfunden und genutzt wird sowie ob an die positiven Auswirkungen von intelligenten Tutorensystemen in vielen anderen Bereichen angeknüpft werden kann [Va06, Sh14].

2 Stand der Forschung

In diesem Abschnitt wird der Stand der Forschung in Bezug auf Unterrichtsentwürfe und computerbasierte Unterrichtsplanungstools dargestellt.

2.1 Unterrichtsentwürfe

Im Allgemeinen besteht ein Unterrichtsentwurf aus vier Abschnitten: Als erstes einem formalen Teil (u. a. Datum, Fach, Raum), gefolgt von einem beschreibenden Teil (u. a. erwartete Lehr-Lern-Ergebnisse, Analyse der Lerngruppe, Begründungen der Lehr-Lern-Situation und -Gegenstands), einer Darstellung des Ablaufs (oft als Tabelle) und Anlagen (z. B. Materialien, Arbeitsblätter, ...) [Pe94, Mü07, Es13]. Dennoch enthält ein Entwurf je nach Ausführlichkeit nicht zwingend immer alle Angaben.

Kriterien guten Unterrichts und Unterrichtsqualität sind Bestandteil zahlreicher Publikationen [Me04, He06, Zi14]: [Me04] und [He06] beantworten die Frage „Was macht guten Unterricht aus?“ mit jeweils zehn Merkmalen, die z. B. *Klarheit und Strukturiertheit*, *Methodenvielfalt* oder *Sicherung* enthalten. Auf der Planungsebene wurden von [KBD15] konkretere Kriterien zur Analyse von schriftlichen Unterrichtsplanungen vorgeschlagen, die auf der Untersuchung basieren, ob bestimmte Aspekte im Entwurf erwähnt bzw. diskutiert werden (z. B. eine Beschreibung der Aufgabenstellung oder Durchführung einer Sicherung) und verbinden damit strukturelle mit inhaltlichen Gesichtspunkten. Ein ähnlicher Ansatz mit einer Bewertungsmatrix wurde von [Ha06] benutzt: Dabei werden neun vorgegebene Aspekte (z. B. Objectives, Activities, Time Allotted, Assessment und Differentiated Instruction), hinsichtlich deren Ausführlichkeit im Entwurf bewertet (z. B. bei Objectives: 0 Punkte für „nicht erwähnt“ bis 4 Punkte für „klar dargestellt und Bezug zu den Standards ist erkennbar“).

2.2 Computerbasierte Unterrichtsplanungssysteme

Zur Unterstützung von Lehrpersonen existiert eine Reihe von computerbasierten Unterrichtsplanungssystemen [SP14, St16]. Diese haben jedoch teilweise sehr unterschiedliche Ziele und Fokussieren auf verschiedene Arten der Unterstützung.

Das wohl erste (publizierte) Unterstützungssystem LPS [Wi00] bildet (optisch exakt) eine papierbasierte Vorlage für Unterrichtsentwürfe in Software ab – durch vorgegebene Eingabefelder soll die kognitive Belastung reduziert werden. Darüber hinaus enthält LPS sowohl weiterführende Informationen (wie z. B. Tipps für gute Pläne) als auch unterstützende Funktionen (Suchen, Drucken, Beispiele ansehen). Darüber hinaus gibt es viele weitere Systeme, die hauptsächlich eine fixe Vorlage zur Eingabe bereitstellen, teilweise aber auch weitere Unterstützungsfunktionen, wie z. B. Rahmenlehrpläne, integriert haben (vgl. [SP14, St16]). Andere Ansätze fokussieren auf eine kollaborative Stundenplanung. Als Beispiel hierfür kann Eduwiki [ZG08] genannt werden, eine wiki-basierte Software zur gemeinschaftlichen Erstellung von Unterrichtsentwürfen mit Funktionen zur Freigabe, Kommentierung, Einladung von Mitarbeitenden und Teilen von Dateien. Ein zentrales Ergebnis einer Umfrage zu Eduwiki war, dass Lehrpersonen Pläne gern freigegeben haben, um selbst Feedback zu erhalten [ZG08]. Zudem gibt es Tools aus dem Bereich Learning Design, wie z. B. LAMS [CC09], die auf eine detaillierte Modellierung des Unterrichtsablaufs fokussieren, um diesen computergestützt (kollaborativ) mit den Lernenden durchzuführen. Eine Besonderheit von LAMS besteht darin, dass mit einem Vorschaumodus bereits während der Planung die Stunde Schritt für Schritt durchgespielt werden kann. Dieser Vorschaumodus wurde in mehreren Studien von den Planenden als besonders hilfreich hervorgehen [CC09].

In einigen Systemen finden sich darüber hinaus auch spezielle Funktionen, die zur Reflexionsanregung der Planung dienen sollen, wie z. B. eine statistische Auswertung der vorgesehenen Sozialformen, Checklisten oder eine graphische Darstellung des Ablaufs (vgl. [CC09, SP14, St16]), jedoch erfordern diese oft meta-kognitive Fähigkeiten um sie „richtig“ einzuschätzen und strategisch zu nutzen (vgl. [Ro11]) – diese Fähigkeiten sind aber gerade bei noch unerfahrenen Lernenden oft noch nicht vorhanden. Eine weitergehende computergestützte, zielgerichtete und proaktive Unterstützung von (angehenden) Lehrpersonen bei der Planung wird jedoch bisher nach bestem Wissen der Autoren von keinem System angeboten; auch wenn dies potenziell viele Vorteile, wie z. B. Adaptivität sowie individuelles und zeitnahes Feedback, bringen kann.

3 Ein Ansatz für automatisches Feedback

Zur Durchführung einer Evaluation wurde ein Ansatz für automatisches Feedback im PLATON-System implementiert [SP14, St16]. Bei PLATON handelt es sich um ein Unterrichtsplanungssystem mit einem graphischen, zeitstrahlbasierten Ansatz (vgl. Abb. 2): Unterrichtsentwürfe entstehen hier nicht in einem reinen Fließtext mit Ablauftabelle, sondern in einer Mischform aus graphischer Ablaufplanung (mit expliziter Modellierung

von z. B. Längen, Sozialformen, erwarteten Lehr-Lern-Ergebnissen, Rahmenlehrplan-bezügen, Ressourcen) und beschreibenden Texten für Erläuterungen/Begründungen.

Die Unterstützung von Lernenden durch Algorithmen basiert grundsätzlich auf zwei Schritten: Zuerst muss eine Analyse durchgeführt werden, um konkrete Aspekte zu finden, bei denen ein Lernender unterstützt werden soll. Diese Analyse findet im hier vorgeschlagenen Ansatz analog zu einem constraint-basierten Tutorensystem [Mi01] über die Überprüfung von Eigenschaften der Lernerlösung statt. Der zweite Schritt beschäftigt sich schließlich damit, den Lernenden Feedback adäquat bereitzustellen.

3.1 Computerbasierte Analyse von Unterrichtsentwürfen

Wie bereits anhand der Kriterien für guten Unterricht bzw. Unterrichtsqualität erkannt werden kann, besteht eine besondere Schwierigkeit bei der Analyse von Unterrichtsplanung darin, dass es sich um eine ill-defined Domäne handelt (vgl. [Ly09, St16]). Es gibt folglich kein „richtig“ oder „falsch“. Dennoch können basierend auf Qualitätskriterien (vgl. Abschnitt 2.1) heuristische Aussagen getroffen werden: Beispielsweise kann überprüft werden, ob eine Beschreibung der Lerngruppe fehlt oder vermutlich zu kurz ist, um alle relevanten Sachverhalte zu beinhalten. Eine solche heuristische Analyse ist jedoch auf Felder mit vorgegebener Semantik beschränkt (z. B. erwartete Lernergebnisse, Lerngruppen- und Aktivitätsbeschreibungen).

- **Stundenthema**
 - nicht angegeben?
- **Rahmenlehrplan**
 - keine Verknüpfung vorgenommen?
 - mehr als drei Aspekte verknüpft?
- **Lerngruppenbeschreibung**
 - nicht vorhanden?
 - zu kurz (<100 Zeichen)?
- **Erwartete Lernergebnisse**
 - nicht vorhanden?
 - zu kurz (<100 Zeichen)?
 - zu viele (>500 Zeichen)?
- **Aktivitäten/Phasen**
 - Beschreibung vorhanden?
 - nur eine Aktivität für die gesamte Stunde?
 - mehr als fünf Phasen?
 - existiert eine Erarbeitungsphase?
 - existiert eine Sicherungs- bzw. Auswertungsphase?
- **Antizipierte Schwierigkeiten**
 - nicht angegeben?
- **Ressourcen**
 - nicht vorhanden?
- **Sozialformen**
 - für alle Aktionen angegeben?
 - überwiegt eine Sozialform (sie nimmt >25 oder sogar >50% einer Stunde ein)?
- **Zeitmanagement**
 - Stunde vollständig verplant bzw. Lücken zwischen den Phasen?
 - Aktivität ist in vorgesehener Zeit durchführbar (Gruppenarbeit oder Unterrichtsgespräch aber <5 Minuten vorgesehen)?
 - Verhältnis der Dauern der Erarbeitungs- und Sicherungsphase ist nicht „optimal“

Abb. 1: Prüfkriterien für Feedbackgabe

Darüber hinaus können bestimmte Strukturen der Planung (z. B. der modellierte Ablauf für eine Analyse herangezogen werden, um mögliche Probleme zu erkennen: z. B. bei der Methodenvielfalt (eine Sozialform überwiegt zeitlich deutlich), bei der Phasierung

(z. B. Sicherung fehlt) oder bei der zeitlichen Planung (z. B. die Unterrichtsstunde ist nicht vollständig verplant oder eine Aktivität ist in der vorgesehenen Zeit vermutlich nicht durchführbar; bspw. Unterrichtsgespräch in 3 min.). Darstellung des generierten Feedbacks

Für die Evaluation wird auf eine Kombination von einfachem visuellem (z. B. Darstellung von grünen Häkchen) und elaboriertem textuellem (z. B. konkreten, beschreibenden Hinweisen) sowie unmittelbarem und on-demand Feedback gesetzt, um die jeweiligen charakteristischen Vor- und Nachteile möglichst optimal zu nutzen [Na06, Va06, Sh08]. Ziel ist es, nicht von der eigentlichen Aufgabe abzulenken oder „erschlagend“ auf die Lernenden zu wirken.

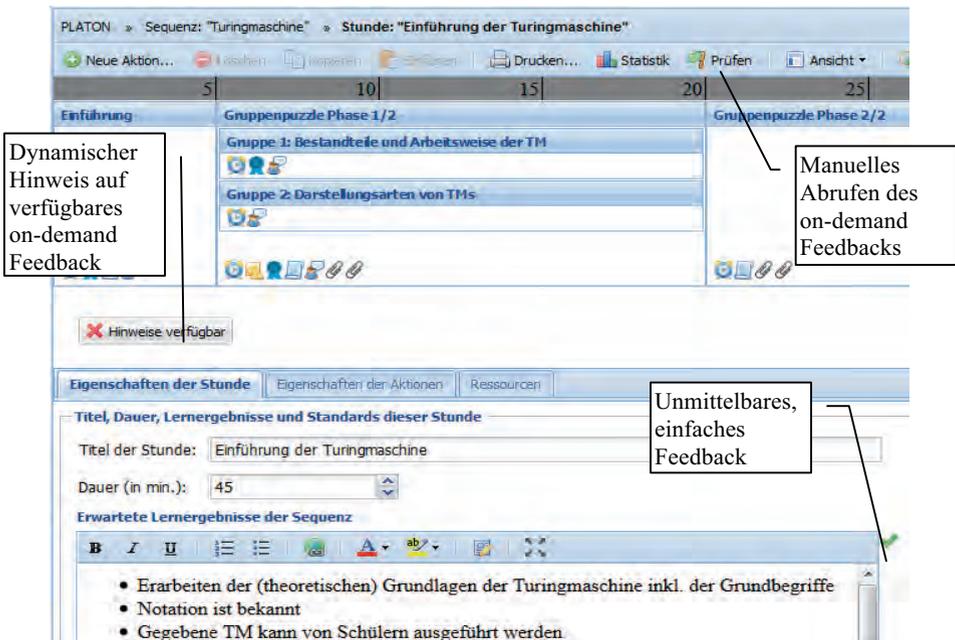


Abb. 2: Planungsansicht einer Unterrichtsstunde im Planungstool PLATON

Bei grundlegenden Feldern, für z. B. die Lerngruppenbeschreibung, wird auf ein einfaches, unmittelbares Feedback in Form von kleinen Häkchen direkt neben den Eingabefeldern gesetzt, die, abhängig von der Eingabelänge, entweder grün oder orange dargestellt werden (vgl. Abb. 2). Ein Tooltip gibt jeweils weiterführende Informationen.

Das on-demand Feedback wird über einen dedizierten sog. „Prüfbericht“ bereitgestellt (vgl. Abb. 3), der von diversen Stellen der Oberfläche über „Prüfen“-Buttons aufgerufen werden kann. Der Dialog ist zweigeteilt: Im oberen Bereich wird die Erfüllung von grundsätzlichen Rahmenbedingungen mit einer Ampeldarstellung visualisiert, um einen schnellen Überblick über etwaige Problembereiche zu ermöglichen. Im unteren Bereich findet sich schließlich eine detaillierte Auflistung von gefundenen möglichen Problemen,

die jeweils „aufgeklappt“ werden können, um weiterführende Informationen in Form

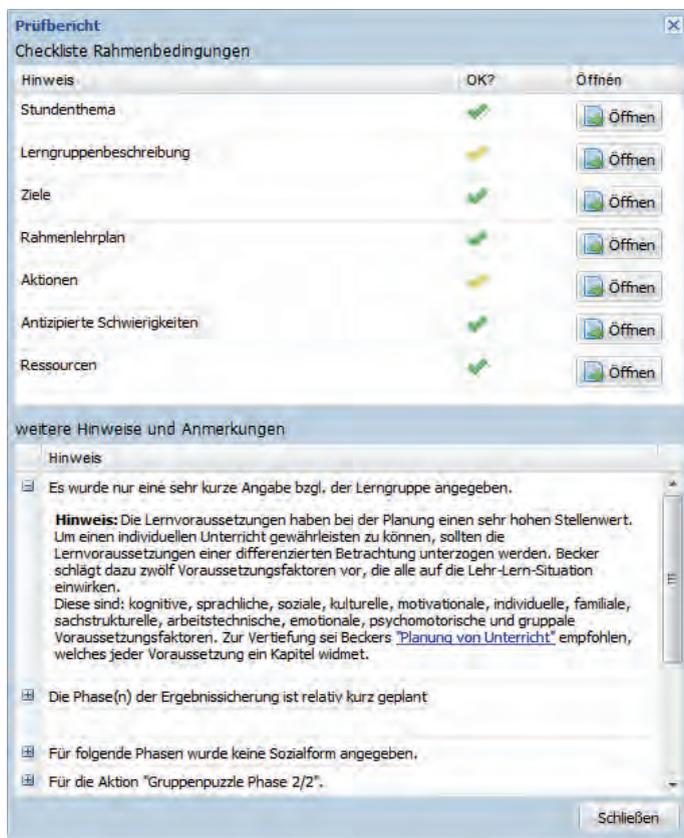


Abb. 3: Darstellung des elaborierten Feedbacks

von Hinweisen, Ratschlägen und Literaturempfehlungen zu erhalten. Da es sich um eine ill-defined Domäne handelt, wurde bei den Formulierungen darauf geachtet, dass möglicherweise unzutreffendes Feedback die Lernenden nicht verunsichert. Da on-demand Feedback prinzipiell die Gefahr birgt nicht abgerufen zu werden, erscheint ein direkt in den Workspace integrierter Button „Hinweise verfügbar“, sofern Feedback verfügbar und die Stunde zu mindestens 70 % verplant ist (Schwellwert zur Vermeidung von Ablenkung durch (noch) unpassendes Feedback, vgl. Abb. 2).

4 Evaluation

Bei der Evaluation des in diesem Paper vorgeschlagenen automatischen Feedbacks standen die folgenden drei Forschungsfragen im Zentrum:

- Wie werden papierbasierte bzw. automatisch generierte Hinweise von Master-Studierenden bewertet und genutzt?
- Inwieweit schlagen sich die Hinweise in der Planung nieder?
- Wie bewerten Master-Studierende die Darstellung des Feedbacks?

Zur Beantwortung dieser Fragen wurde eine dreistündige Laborstudie mit einem between-subjects Design durchgeführt: Die teilnehmenden Personen wurden dazu in zwei Gruppen eingeteilt, wobei die erste Gruppe (Kontrollgruppe, „Gruppe P“) einen ausgedruckten Papier-Leitfaden mit allgemeinen, ausformulierten Hinweisen für einen Unterrichtsentwurf (z. B. wie und warum Lernvoraussetzungen bzw. Lernziele ausformuliert werden sollten) erhielt. Die zweite Gruppe („Gruppe A“) konnte die in PLATON implementierten Feedback-Funktionen nutzen. Im Grunde decken sich die ausformulierten Hinweise mit den Aspekten, die das automatische Feedback überprüft.

Der Ablauf war grundsätzlich für beide Gruppen identisch und bestand aus einer Vorstellung des Systems PLATON (ca. 15 min.) mit Hilfe eines Videos (identisch für beide Gruppen – für Gruppe P fehlte lediglich die Feedback-Funktion), gefolgt von einer ca. 110-minütigen Arbeitsphase, in der eine 90-minütige Unterrichtsstunde zu planen und in eine grobe Unterrichtsreihe (von 4-6 Stunden) einzuordnen war. Eine Lerngruppenbeschreibung sowie das zu planende Fach wurden vorgegeben. Das Thema der Stunde durften sich die Teilnehmenden anhand des Rahmenlehrplanes selbst wählen. Abschließend erfolgten die Bearbeitung eines Fragebogens mit einem offenen und geschlossenen Teil (ca. 15 min.) zur Abfrage der Nutzung, Qualität sowie Limitationen der Hinweise und Verbesserungsvorschlägen.

Während der Studie wurden die Bildschirminhalte der teilnehmenden Personen aufgenommen und sämtliche Interaktionen mit dem System protokolliert. Damit konnte im Anschluss an die Studie ermittelt werden, wie oft die Feedback- bzw. Prüfen-Funktionen genutzt und welche Änderungen im Anschluss an der Planung vorgenommen wurden.

Die Teilnahme an der Studie stand grundsätzlich allen Master-Lehramtsstudierenden in Berlin mit guten Deutsch-Kenntnissen sowie eigener Unterrichts- bzw. Planungserfahrung (mindestens ein Unterrichtspraktikum) offen und wurde mit 30 € honoriert. An der Studie haben insgesamt sechzehn Master-Studierende (6 w. u. 10 m.) der Humboldt-Universität zu Berlin teilgenommen; sieben Studierende in der ersten und die übrigen neun in der zweiten Gruppe. Die Fächer der Studierenden deckten sowohl alle Haupt- als auch diverse natur- und geisteswissenschaftliche Fächer sowie Sprachen und Sport ab.

5 Ergebnisse der Fallstudie

Alle Teilnehmenden waren in der Lage, Unterricht mit Hilfe von PLATON zu planen. Jedoch gibt es in der Gruppe A zwei Personen (A-5 und A-8), die sich sehr intensiv mit der Recherche und Sachanalyse beschäftigt und den Ablauf der Stunde nur grob geplant haben. Zudem hat eine Person (P-1) die Unterrichtsstunde nicht vollständig geplant.

5.1 Nutzung der Hinweise

Als erstes wurden die am Ende der Studie vorliegenden Entwürfe dahingehend untersucht, wie viele Hinweise das Feedbacksystem auf Basis dieser finalen Entwürfe noch geben konnte (per Studiendesign stand das System nur der Gruppe A während der Bearbeitung zur Verfügung). Anhand Abb. 4 ist zu erkennen, dass für die Gruppe P insgesamt 28 Hinweisen deutlich mehr generiert werden, als für die Gruppe A mit 21 Hinweisen. Hierbei sind jedoch zwei wichtige Aspekte zu beachten: 1) Die Gruppe P besteht aus sieben und die Gruppe A aus neun Personen. 2) Da es sich um die finalen Entwürfe handelt, sind bei der Gruppe A alle Hinweise, die bereits während der Bearbeitung berücksichtigt wurden, nicht mehr enthalten. Dennoch ist dieser Unterschied zwischen den beiden Gruppen gemäß U-Test bei einem Signifikanzniveau von 5 % nicht signifikant ($p=0,071$). In der Gruppe A waren zwei Personen so sehr mit der Planung beschäftigt, dass sie das generierte Feedback nicht zur Kenntnis genommen haben, und diese beiden Personen vier bzw. acht Hinweise erhalten hätten. Die meisten Hinweise bezogen sich dabei auf Rahmenbedingungen, wie z. B. Verknüpfungen mit dem Rahmenlehrplan, Lerngruppenbeschreibungen und antizipierte Schwierigkeiten.

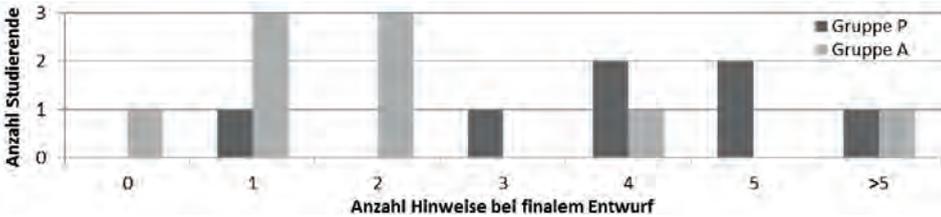


Abb. 4: Häufigkeitsverteilung der Feedbackhinweise des finalen Entwurfs

Basierend auf Beobachtungen während der Unterrichtsplanung und Antworten auf dem Fragebogen hat die Kontrollgruppe P den ausgedruckten Hinweisen insgesamt nur wenig Beachtung geschenkt und diese ggf. überflogen. Das automatische Feedback hingegen wurde von vielen direkt bei der Planung beachtet: Der Prüfbericht wurde von fast allen Studierenden mit Ausnahme von A-4, A-5 und A-8 während der Planung abgerufen. Darüber hinaus wurde der Prüfbericht von den Personen A-2 (10x), A-3 (17x) und A-7 (12x) sehr intensiv genutzt – im Durchschnitt fünfmal. Dies betrifft jedoch nur den aktiv abgerufenen Prüfbericht; über das unmittelbare Feedback bei den Eingabefeldern sind hiermit keine Aussagen möglich. Das Feedback bezog sich vor allem auf die Lernvoraussetzungen, Rahmenlehrplan-Verknüpfung und antizipierte Schwierigkeiten. Insgesamt haben sechs Personen der Gruppe A (67 %) maximal zwei Hinweise nicht in ihrer Planung berücksichtigt. 71 % der Teilnehmenden nahmen direkt nach Abruf des Feedbacks Änderungen vor. Dies betraf vor allem die Verknüpfung des Rahmenlehrplans (5x), die Überarbeitung der Lerngruppenbeschreibung (4x), das Hinzufügen von antizipierten Schwierigkeiten (2x) sowie eine Überarbeitung des Ablaufs (1x).

Vor allem die Lerngruppenbeschreibungen sind besonders erwähnenswert: In der Gruppe A wurde für keinen Teilnehmenden nach der Studie ein Hinweis hierfür

generiert. Fünf Personen haben diese im Rahmen der Planung überarbeitet und eine Person (A-2) ersetzte ihre aus Stichpunkten bestehende Beschreibung durch einen ausführlicheren Fließtext. Die insgesamt aufgewendeten Bearbeitungsdauern sind bei beiden Gruppen in etwa identisch (Gruppe P: $d=6$ min; $sd=3,07$; Gruppe A: $d=8$ min; $sd=3,7$; $p=0,152$ gemäß U-Test). Jedoch gibt es einen statistisch signifikanten Unterschied in der Länge (Gruppe P: $d=380$ Zeichen; Gruppe A: $d=885$ Zeichen) der Beschreibungen ($p=0,031$ gemäß U-Test) und in der Anzahl von Bearbeitungen (Gruppe P: $d=1$; Gruppe A: $d=1,75$; $p=0,04$ gemäß U-Test), wobei beide Werte bei Gruppe A höher sind. Bei den erwarteten Lernergebnissen für die zu planende Unterrichtsreihe sowie bei den antizipierten Schwierigkeiten sind keine signifikanten Unterschiede feststellbar. Auffällig ist jedoch, dass A-8 als einzige Person keine erwarteten Lernergebnisse bei der Unterrichtsreihe angegeben hat.

5.2 Auswertung der Fragebögen

Im Folgenden werden die Ergebnisse der Fragebögen überblicksartig zusammengefasst: Die Bewertung erfolgte auf einer 5-stufigen Likert-Skala mit den Extrema „stimme gar nicht zu“ und „stimme voll zu“, die zur Auswertung auf die Werte 1 resp. 5 abgebildet wurden. In der Gruppe P gibt es eine leichte Zustimmung, dass allgemeine Hinweise ausreichend seien ($m=4$). Die vorliegenden Hinweise wurden zudem zumeist als hilfreich angesehen ($m=4$). Dennoch waren die Empfindungen bzgl. Reflexion sowie Beeinflussung direkt bei der Planung durch die Hinweise sehr breit gestreut (jeweils $m=3$) und es herrschte eher die Meinung, dass die allgemeinen Hinweise auf keine neuen Aspekte aufmerksam gemacht haben ($m=2$). Dazu passt, dass die Studierenden der Aussage „Ich habe meine Planung aufgrund der gegebenen Hinweise angepasst bzw. überarbeitet.“ eher nicht zustimmten ($m=2$). Insgesamt waren sich die Teilnehmenden der Gruppe P unsicher, ob PLATON ebenfalls diese Hinweise geben könne ($m=3$). Trotzdem wurden konkrete Hinweise durch ein System im Vergleich zu den allgemeinen Hinweisen tendenziell als effektiver angesehen ($m=4$), die auch solche allgemeinen Hinweise ersetzen könnten ($m=4$). Von der Gruppe A wurde automatisches Feedback tendenziell als individuell ($m=4$), recht hilfreich ($m=2$) und angemessen ($m=3$) empfunden. Dennoch herrschte relativ große Einigkeit darin, dass das automatische Feedback Tutorenfeedback nicht ersetzen könne ($m=1$). Die Mehrzahl der Studierenden gab an, ihre Planungen auf Grund des erhaltenen Feedbacks überarbeitet zu haben ($m=4$) und durch die Hinweise zur Reflexion angeregt worden zu sein ($m=4$). Trotzdem sah, mit Ausnahme einer Person (A-7), die Mehrheit die Hinweise als tendenziell zu allgemein an ($m=3$). Obwohl sich viele Studierende scheinbar keine Aussage zur didaktischen Qualität und Frage, ob die automatischen Hinweise zu oberflächlich waren, zutrauten und die mittlere Bewertung wählten, stimmten viele zu, dass sie die Hinweise auf „neue“ Aspekte aufmerksam gemacht haben ($m=4$). Insgesamt wurde die Darstellung und Gestaltung des Feedbacks/Prüfberichts als gelungen empfunden ($m=4$).

Bei der Gruppe A empfanden zwei Studierende, dass „nur formale Aspekte“ überprüft werden und die Länge einer Lerngruppenbeschreibung nichts über deren Qualität aussagt

(A-6), als Einschränkung. Dennoch wurde angemerkt, dass die generierten Hinweise viel konkreter und „besser“ als sowohl ein allgemeiner Leitfaden (A-2 und A-3) als auch eine eigene Überprüfung seien. Einige Studierende sahen positiv, dass man automatisches Feedback nicht so leicht übersehen könne – bei der Gruppe P hingegen gab über die Hälfte der Studierenden an, die allgemeinen Hinweise nicht gelesen oder lediglich überflogen zu haben. Dies deckt sich mit Aussagen wie z. B. „Hilfe für Unerfahrene (...) [, aber] sind wenig individuell und helfen evtl. nicht[,] die Planung zu verbessern, denn grundsätzl. sollten die Eckdaten dafür angehenden Pädagogen bekannt sein“ (P-6) und „In Textform ohne Aufzählungen oder Stichpunkte wirken sie häufig erschlagend“ (P-4). Konkret wurden die automatischen Hinweise zu methodischen Überlegungen (A-3), zu möglicherweise zu kurzen Phasen (A-7) und zu antizipierten Schwierigkeiten (mehrfach) als besonders hilfreich bezeichnet.

Darüber hinaus gab es Verbesserungsvorschläge: Person A-8 nutze die Hinweise nicht und schlug vor, nach einer gewissen Zeit auf die Prüfen-Funktion explizit hinzuweisen. Andererseits sollte das Feedback auch punktuell deaktiviert werden können, wenn z. B. bewusst keine Lernvoraussetzungen angegeben wurden (A-8).

6 Diskussion

Als Startpunkt der Evaluation wurde die Anzahl der generierten Hinweise der beiden Gruppen für die endgültige Planung verglichen. Da designbedingt die Hinweise auf Qualitätskriterien basieren und vor der Studie ausgiebig getestet wurden, können sie als ein Indikator für die Qualität gesehen werden (vgl. Abschnitt 2.1). Eine komplette inhaltliche Analyse der Entwürfe war im Rahmen dieser Studie leider nicht möglich.

Insgesamt schenkte die Gruppe A den Hinweisen deutlich mehr Aufmerksamkeit als die Gruppe P – obwohl die vorbereiteten Hinweise eine sehr große Übereinstimmung aufwiesen: Die Beachtung der papiergebundenen allgemeinen Hinweise liegt im Verantwortungsbereich der Studierenden und müssen selbständig überprüft werden. Einige haben sie kurz überflogen, intensiv gelesen wurden sie nicht – sie seien „schon bekannt“, hätten nicht bei der Verbesserung der Planung oder wirkten „zu erschlagend“. Das generierte Feedback hingegen basiert auf einer automatischen Prüfung und wird aktiv bereitgestellt. Zudem enthielt der Prüfbericht eine deutlich kürzere, adaptiv für jeden Nutzenden gefilterte Liste, die (sehr wahrscheinlich) zutreffende Aspekte betreffen. Dies erhöht sicherlich die wahrgenommene Sichtbarkeit einzelner Hinweise. Dennoch ist auffällig, dass die mehrheitlich präsentierten Hinweise zu formalen Punkten (z. B. Rahmenlehrplan und Lerngruppenbeschreibungen) tendenziell als sehr allgemein und weniger relevant empfunden wurden, aber trotzdem oft zu einer Überarbeitung führten.

Auffällig ist, dass einige Personen den Prüfbericht sehr intensiv genutzt ($\max=17$; $d=5$) und versucht haben, die dargestellten Hinweise „abzuarbeiten“. Diese extensive Nutzung könnte auf das „Gaming the system“-Phänomen hindeuten [Ba08]. Im Rahmen der Studie wurden jedoch nur sinnvolle Änderungen beobachtet, wobei hier nicht bewertet

werden konnte, ob alle ergänzten Details für die konkrete Planung wirklich erforderlich waren. Dennoch haben einige Studierende absichtlich Hinweise ignoriert, woran man erkennen kann, dass es ihnen sicherlich nicht nur darum ging, im Prüfbericht nur „grüne Haken“ vorzufinden. Korreliert man dies noch mit den Fragebögen, erkennt man, dass die Hinweise von vielen als hilfreich und als Anregung zur Reflexion eingestuft wurden.

7 Zusammenfassung und Ausblick

In diesem Paper wurde ein erster Ansatz für automatisches Feedback für angehende Lehrpersonen zur Unterstützung der Unterrichtsplanung vorgestellt. Insgesamt handelt es sich um einen ersten Schritt für regelbasiertes, algorithmisch generiertes Feedback in dieser Domäne. Dabei hat sich in einem Vergleich zu papierbasierten Leitfäden das automatische adaptive Feedback zu Mindestanforderungen und Planungsauffälligkeiten als wirkungsvoller hinsichtlich der Beachtung und Anregung zur Reflexion erwiesen.

Sicherlich kann das derzeit verfügbare automatische Feedback keine menschlichen Tutoren ersetzen, die auch konkretes inhaltliches Feedback geben können. Dennoch kann es, über die aktuell geprüften formalen Aspekte hinaus, weiter ausgebaut werden: Insbesondere bei kompetenzorientierter Planung wird die Nutzung sog. Operatoren (z. B. „erläutern“, „bewerten“) zur Beschreibung gefordert. Daher könnten diese, deren Fehlen oder Nutzung von nicht messbaren Formulierungen wie „Die SuS wissen...“ ebenfalls für weiteres Feedback herangezogen werden. Zudem waren Master-Studierende in dieser Studie die Zielgruppe. Fortgeschrittenere Lehrpersonen benötigen wahrscheinlich eine andere Art von Feedback, daher könnte das Feedback auch an die Planenden adaptiert werden, so dass z. B. Hinweise, die mehrfach nicht beachtet wurden, nicht erneut angezeigt werden, oder mit der Zeit weniger bzw. tiefergehendes Feedback präsentiert wird. Darüber hinaus könnten, im Gegensatz zu festen Regeln, Parameter guter Entwürfe auch maschinell gelernt werden.

Danksagung

Ein besonderer Dank gilt Daniela Pirschel, die im Rahmen einer durch die Autoren dieses Papers betreuten Masterarbeit die Feedback-Komponente im PLATON -System entwickelt und den Grundstein der Evaluation gelegt hat.

Literaturverzeichnis

- [Ba08] Baker, Ryan et al.: Why students engage in "gaming the system" behavior in interactive learning environments. JILR 19/2, S. 185-224, 2008.
- [CC09] Campbell, Chris; Cameron, Leanne: Using Learning Activity Management Systems (LAMS) with pre-service secondary teachers: An authentic task. In Proc. ascilite, 2009.

- [Es13] Esslinger-Hinz, Ilona et al.: Der ausführliche Unterrichtsentwurf. Beltz, 2013.
- [Ha06] Hansen, Cory C.: Technology as an electronic mentor: Scaffolding preservice teachers in writing effective literacy lesson plans. *NAECTE JECTE*, 27/2, S. 129-148, 2006.
- [He06] Helmke, A.: Was wissen wir über guten Unterricht? *Pädagogik* 58/2, S. 42-45, 2006.
- [KBD15] König, Johannes; Buchholtz, Christiane; Dohmen, Dieter: Analyse von schriftlichen Unterrichtsplanungen. *Z Erziehungswiss* 18/2, S. 375-404, 2015.
- [Ly09] Lynch, Collin et al.: Concepts, structures, and goals: Redefining ill-definedness. *IJAIED* 19/3, S. 253-266, 2009.
- [Me04] Meyer, Hilbert: Was ist guter Unterricht? Cornelsen Scriptor, 2004.
- [Mi01] Mitrovic, Antonija et al.: Constraint-based tutors: a success story. In: *Proc. IEA/AIE 2001*, S. 931-940, 2001.
- [Mü07] Mühlhausen, Ulf: Unterrichtsvorbereitung – wie am besten? In (Daschner, Peter; Drews, Ursula, Hrsg.): *Kursbuch Referendariat*, 2007.
- [Na06] Narciss, Susanne: Informatives tutorielles Feedback: Entwicklungs- und Evaluationsprinzipien auf der Basis instruktionspsychologischer Erkenntnisse. Waxmann, 2006.
- [Pe94] Peterßen, Wilhelm H.: *Handbuch Unterrichtsplanung*. München: Ehrenwirth, 1994.
- [Ro11] Roll, Ido et al.: Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction* 21/2, S. 267-280, 2011.
- [Sh14] Sharma, Shanky et al.: Survey of Intelligent Tutoring Systems. *IJERT* 3/11, 2014.
- [Sh08] Shute, V. J.: Focus on Formative Feedback. *Rev. Educ. Res.* 78/1, S. 153-189, 2008.
- [SP14] Strickroth, Sven; Pinkwart, Niels: Softwaresupport für die graphische, zeitbasierte Planung von Unterrichtseinheiten. In *Proc. DeLFI 2014*, S. 314-319, 2014.
- [St16] Strickroth, Sven: Unterstützungsmöglichkeiten für die computerbasierte Planung von Unterricht. Dissertation. Humboldt-Universität zu Berlin, 2016.
- [Va06] VanLehn, Kurt: The behavior of tutoring systems. *IJAIED* 16/3, S. 227-265, 2006.
- [Wi00] Wild, Martyn. Designing and evaluating an educational performance support system. *British Journal of Educational Technology*, 31/1, S. 5-20, 2000.
- [ZG08] Zhou, Yueliang; Gong, Chaohua: Research on Application of Wiki-based Collaborative Lesson-preparing. In *IEEE WiCOM'08*, S. 1-5, 2008.
- [Zi14] Zierer, Klaus: *Hattie für gestresste Lehrer*. Baltmannsweiler: Schneider Verlag, 2014.

Analytics Prototype for Data Driven Decision Making for Blended Learning Strategies in HEI

Vlatko Lukarov¹, Ulrik Schroeder¹

Abstract: In this paper we present a tool that introduces data-driven decision making concerning the implementation and adoption of blended learning in higher education. We identified key groups from the university's administration that are directly involved in the learning and teaching processes and captured their goals and perspectives in the context of eLearning processes. We used these perspectives and goals to sketch out systematical introduction of data-driven decision making processes which can create sustainable impact on their work and the blended learning scenarios. This resulted in building a web-based prototype that visualizes usage statistics and analytics of the log data extracted from the university-wide learning platform. This prototype contains indicators and visualizations mapped to the goals and perspectives of the identified key user groups, in order to provide insight how the different faculties teach and learn utilizing the platform and provides actionable intelligence to better distribute resources and support the staff and students in their respective blended learning scenarios.

Keywords: Decision Making, Analytics, Learning Data, blended learning strategy

1 Introduction

Evidence based, or data based decision making is the process of analyzing and evaluating data to improve learning offerings and resources, curriculum, and make informed decisions about the teaching and learning processes [Cr06] [Ho12]. These institutional decisions are too important and critical to be based on intuitions, or presumptions. These decisions require facts, knowledge, and analytics. In theory, visual analytics and data analytics should be the main sources of information in decision making, representing an improvement over intuition. Unfortunately, in practice, this is not the case. Often, decision making is usually based on intuition, presumption, and on accumulated experience, without any specific data or analysis [CDO07]. Since, most universities and higher education institutions (HEI) in Germany are public institutions, which means there is a certain level of transparency and accountability to the public not only of the decisions, but also of the decision-making process. Hence, such position affords the higher education institutions to evaluate, revise, and improve the tools they use for decision making.

When it comes to institutional decisions and decision-making in higher education institutions concerning the learning and teaching processes, there are two main groups of stakeholders: students and faculty, and administrative bodies. In the second group, there

¹ Learning Technologies, Informatik 9, RWTH Aachen University Ahornstraße 55, 52074 Aachen, Germany, {Lukarov, schroeder}@informatik.rwth-aachen.de

are students and faculty members that belong to different administrative bodies, but in this case, they have different roles[Ch12]. The ongoing work presented in this paper focuses on the administration stakeholder group. This group includes the blended learning coordinators² (one for each faculty), the committee for eLearning (consisting of all BL coordinators, plus the management leadership from the institutes responsible for eLearning processes and implementation), the rectorate and dean(s) responsible for teaching, academic affairs, planning, development, control and institutional research, and the center for teaching and learning (including the developers' team). Each of these users (or user groups) has different responsibilities, perspectives, and goals concerning their eLearning process and scenarios. In the following section, we will outline the salient points of the goals and perspectives of the different people within the administration. This work is part of an ongoing project at RWTH Aachen University, called AIX – Future teaching & learning funded by Stifterverband which aims to provide the HEI administration possibilities to examine, classify and evaluate their investment and efforts in fostering blended learning concepts and implementations.

2 Goals and Expectations

In the previous section, we presented several user groups who are directly involved in the eLearning processes and infrastructure at RWTH Aachen university. The goals and perspectives from the different user groups were collected in two phases. In the first phase, we conducted a literature review and several brainstorming sessions to identify and sketch out the different perspectives of the different user groups. The second phase were a series of semi-structured interviews with different people from the identified users' groups to gain an insight about how they collect and gather pieces of information, and how they use this information in their decision-making process. The intermediate results from the first phase were used as a basis for guidance of the semi-structured interviews in the second phase. The different involved groups and the salient points of their respective goals and perspectives are summarized in Table 1.

It is noteworthy to mention that different groups responsible for different aspects of the eLearning processes have similar, or even common goals, but still have different perspectives. For example, the developers' team is interested to learn how different faculties use the various functionalities of the learning platform in their course rooms and how (when, how much, which devices) the users interact with the different modules in these courses. This way they can evaluate their modules, identify patterns of usage, potential problems, system load, and other parameters that could help improve the platform from a developer's/technical point of view. At the same time, the blended learning coordinators, who are responsible for implementation of the blended learning initiative would like to know the same thing, but from a different perspective. They also

² Blended Learning Coordinators are responsible for implementing the blended learning initiative at RWTH Aachen University. They are part of a rectorate appointed committee whose goal is to oversee the university-wide implementation of blended learning concepts.

need this information, how different users from different faculties use the learning platform; on which modules, they use to implement their teaching and learning approach; does their approach relies on lecture scripts, or also includes videos and media; whether collaboration and discussion is useful; is formative assessment integral part of their courses; and many questions that follow under implementation of different eLearning approaches and strategies.

Another example is that the members of the department for planning, development, and control is interested in how different faculties use the underlying learning and technical infrastructure. They need to have an overview and compare how different faculties implement their eLearning initiative(s), which also coincides with the perspective of the coordinators, who need to oversee and analyze that on individual faculties to come to different conclusions and decide upon different actions and measures.

User group(s)	Goals and Perspectives
Developers Team IT Staff Platform Support and Qualification Team	<ul style="list-style-type: none"> • Learning Platform Usage Patterns • Concurrent users and requests (traffic) • Different modules use frequency • OS/Devices distribution among users • Errors and Bugs identification
(Blended Learning) eLearning Committee Blended Learning Coordinators	<ul style="list-style-type: none"> • Resources planning • Implementation of different eLearning approaches • Feedback and guidance for faculty and students <ul style="list-style-type: none"> ○ Continuous Learning vs. Cramming • Success rate/Evaluation of Blended Learning initiative • Scientific literature/library resources use
Rectorate Department for planning, development, and control	<ul style="list-style-type: none"> • Accountability and University Image • Resource Allocation • Effective use of learning resources and infrastructure among different faculties

Table 1. Users involved in decision making with their respective goals and perspectives

Another aspect that was revealed to us during the interviews was that to gather and aggregate information about how to analyze and decide over, is the lack of tools and lack of readily-available and understandable data, which can support them during their work. They do have different data available from them through different channels (mostly qualitative) such as: number of (new) students, students evaluating teaching, log data, interviews, and surveys, but nothing conclusive nor systematic on a regular basis. Furthermore, they must invest a lot of time and resources just to understand the data, because it comes in a variety of forms, factors, and amount, and thus requires a versatile

skill set which goes beyond the scope of their work. This makes it very challenging to analyze and assess the eLearning scenarios and situations just to get an overview of the situation, and nearly impossible to make informed decisions and undertake measurements to achieve their respective goals regarding the eLearning initiatives and activities at RWTH Aachen University.

3 Project

In the previous sections, we presented the basis why the administration needs an analytics tool to support them in achieving their goals in the context of eLearning. According to [Ko08], in order to systematically introduce evidence based or (data-driven) decision making, there is a four-phase process that can help in creating sustainable impact on the learning/teaching processes. The four phases are: data collection, data connection (connecting the data), working with the data, and findings confirmation. We followed this process methodology when we worked on the AiX Analytics prototype. The prototype itself is a web-based prototype that visualizes the usage statistics and analytics of the log data from the learning platform at RWTH Aachen University. The prototype consists of three main building blocks: data analysis and management, RESTful application engine, and User Interface. The building blocks correspond with the first three parts of the process methodology.

3.1 Data Management and Analysis

The primary data source of our prototype is the anonymized log data generated from the learning platform L²P used at RWTH Aachen University. We receive the anonymized version of the log data from the servers that run the learning platform on regular basis for the past year. The data itself contains the timestamp, hashed IP address of the request, the client agent, the processing time, the request/response size, the URI of each request, and the response code. The platform logs that we receive are raw and unfiltered. This means that before we start the analysis, we need to clean up the data and prepare it for analysis. After this step, the analyzers extract and aggregate the information, and save the derived data and useful information. The raw data is kept for 15 days and deleted (data privacy conformant).

3.2 Data Connection

We built a RESTful application engine to connect the derived and analytics data with the user interface. The RESTful application engine is a Web API application consisting of different web services methods (HTTP services) that provide data and analytics to the user interface. The HTTP methods deliver the requested results in JSON format. The application engine itself is modular and already has wide range of methods, and, if necessary, it is straightforward to extend it with new methods.

3.3 Working with Data (User Interface)

The user interface provides a set of visualizations that presents the learning data in various forms interactively for the different users' scenarios and goals. The main emphasis of the visualizations in the prototype is easy-to-understand visualizations, which the stakeholders can use to support and enhance their day to day activities, which can in turn help in achieving their respective goals. We split the interface into two main aspects. The first aspect covers the platform itself and the different modules that exist in each course room. The second aspect covers individual faculty's overall activities divided per semester and summarized over the entire timespan of the available data. The visualizations are interactive (filtering and zoom-in functionality) and enable the user to focus on specific parts of the visualization. The reason behind this split can be inferred from the goals and perspectives of the different stakeholders. Although the users have similar, or overlapping goals, they have different perspectives. We used knowledge and guidelines for implementation of analytics tools and prototypes, which we have already gathered from previous experiences and research. In [Dy14] is provided a requirements catalogue which depicts in details what one needs to consider when building analytics prototypes and tools. We selected a subset of these requirements and guidelines according to the questions we need to answer, the data analytics literacy of our user groups, and the types of visualizations we needed to develop. The visualizations in the prototype are simple and responsive, thus the cognitive effort upon the user is low, and she can concentrate on the visualizations themselves, instead of concentrating on understanding the interface. Furthermore, the visualizations themselves are interactive (filtering and zoom-in functionality) which enables the user to focus on specific parts of the visualization and drill-down into the data to get deeper understanding about the data. The interactions with the visualizations are always coupled with informative tooltips (bound to mouse-hover events) to provide feedback about what is displayed at a given time and place [Dy14]. We developed around 30 different indicators, which try to answer the questions and goals we extracted in our phase. For the visualizations themselves, we use the Highcharts³, interactive JavaScript chart library. In the next section we will display three scenarios /use-cases which show the tool's potential, and also provide interesting insights about how different faculties and field of study implement the blended learning scenarios in HEI.

4 Use Cases and Insights

The first scenario is to see how different faculties use the assessment functionalities on the learning platform. This scenario is useful for the blended learning coordinators and the developers' team. The coordinators can gain an overview of how the different faculties implement formative assessment in their eLearning scenarios. The developers' team can identify different usage patterns of the assessment features on the platform in order to provide more streamlined learning and teaching experience. Figure 1 shows the usage over

³ <http://www.highcharts.com/>

one year of the assessment modules available on the learning platform (Assignments, eTests, Gradebook, and Exam Results).

As one can see that during the semesters (from April till July 2016, and October 2016 till February 2017) the largest and continuous usage of different assessment modules comes

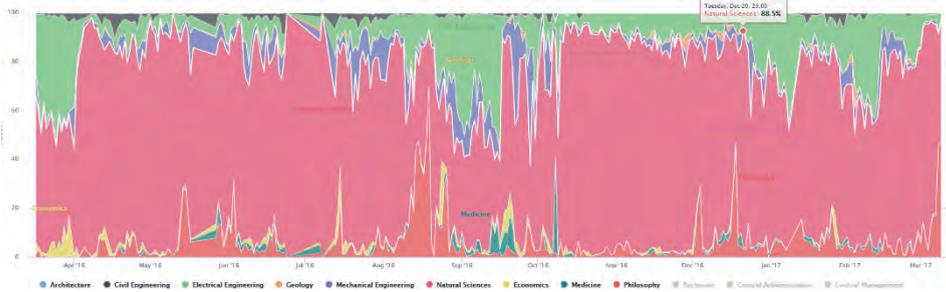


Figure 1 Assessment Distribution among the faculties

from the Faculty of Mathematics, Informatics, and Natural Sciences (Faculty 1). In the exam phases, the usage of assessment modules increases because other faculties publish the results of the exams. From this one can conclude that Faculty 1 uses the assessment modules from the learning platform for formative assessment as an integral part of their blended learning strategies. The other faculties (for example Civil Engineering, Geology) are lagging behind. This can be seen from Figure 2. Figure 2 shows the usage of assignments' module at the Faculty of Mathematics, Informatics, and Natural Sciences, and usage of assignments at the Faculty of Civil Engineering. One can clearly identify the weekly peaks on the left which correspond to weekly assignments, while on the right there

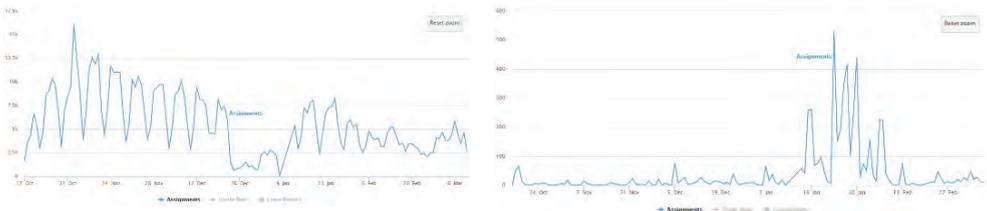


Figure 2 Assignments on Faculty of Natural Sciences vs. Civil Engineering

are only peaks for two weeks in January.

The second scenario involves usage of different media in the learning processes and scenarios at the different faculties. If we look at the overall media use, distributed among the faculties, half of the usage comes from the Faculty of Mechanical Engineering (Figure 3). This visualization is intended for the eLearning committee and the department for planning, development, and control. These stakeholders can get an overview which faculties implement video-based learning, and how it is distributed among the different faculties.

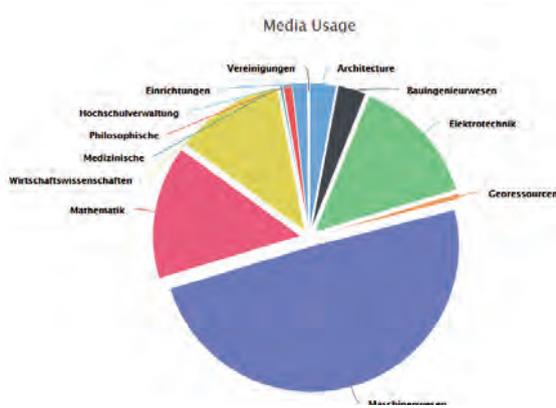


Figure 3 Overall media use among the faculties

On the other hand, the usage numbers are not normalized and correlated with number of students for each faculty, these numbers are understandable and expected. The Faculty of Mechanical Engineering is the biggest faculty at RWTH Aachen University. Furthermore, what was surprising for us was that the Faculty of Architecture relies on media resources. As it can be seen from Figure 4, one third of their usage falls under media usage. After this follow the faculties for Mechanical Engineering, Electrical Engineering, Economics,

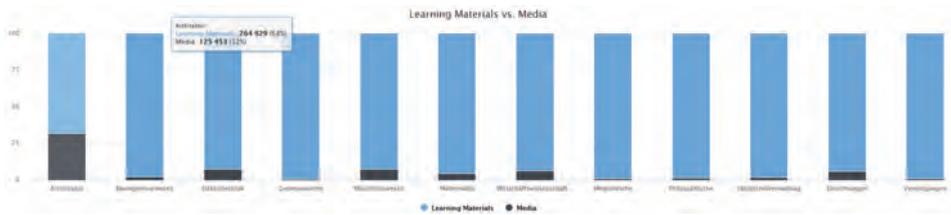


Figure 4 Learning Materials vs. Media Resources per faculty

and Mathematics, Informatics, and Natural Sciences. Their media usage ranges from five to eight percent in comparison with “classic” learning materials. The faculties for Civil Engineering, Geology, Philosophy, and Medicine fall further behind with only one percent. This information could be used to investigate in more details how each faculty implements its video-based learning, and thus allocate more resources and support in their direction.

The third scenario is more general and shows how many users and requests the learning platform has on daily basis; with what kind of devices the users use the learning platform; and which parts of the learning platform are used over one year. This visualization is important for the IT Staff to better scale the hardware, for the rectorate in order to grasp just how many different users rely on the learning platform on daily basis. Figure 5⁴

⁴ The troughs which go to zero in the visualization are points in time with missing log data.

represents a logarithmic representation of the daily requests and unique clients that there are on the platform every day. On weekdays, there are from 16 – 22 000 unique clients (users), and 1.5 – 2.5 million requests, while on weekends the numbers drop to 8-10 000 unique clients, and 0.5 – 1.2 million requests. This overview of unique clients and requests we explored it into two directions, namely the distribution of different devices and the amount of usage of different parts of the platform.

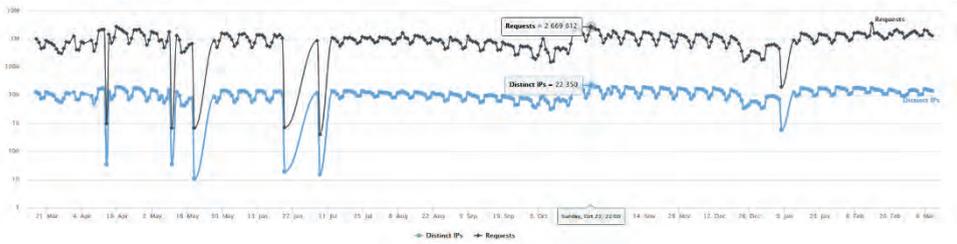


Figure 5 Unique clients and requests on the platform

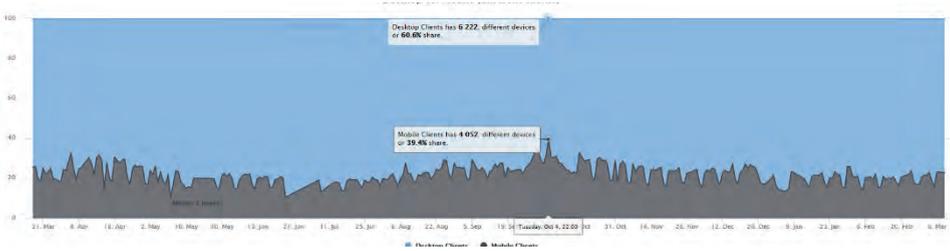


Figure 6 Desktop and Mobile devices on the platform

As it can be seen from Figure 6, on regular basis around 25 percent of the use comes from mobile devices. At the beginning of every semester there are peaks (sometimes around 30-40% of the devices are phones, or tablets) which can be explained with the fact that the students are looking and registering for courses and lectures, and are trying to organize their course work for the semester. This visualization is significant for the development team, because with it they can analyze the usage behavior based on different devices in order to provide better mobile compatibility and support of the platform’s interface.

The course room structure of the learning platform consists of six domains: Organization,

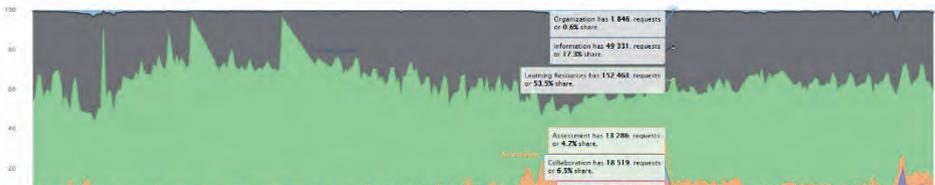


Figure 7 Course room domains use on the learning platform

Information, Learning Resources, Assessment, Collaboration, and Management. Each domain contains modules connected to its domain. Figure 8 shows an aggregated view of how these six domains of the course were used. The two domains with highest usage are the learning resources and information distribution, amounting to 70-80 percent of the use. They are followed by the collaboration and assessment domains, which each see usage ranging from four to ten percent each. One can further drill down in these domains, too see the distribution, participation, and presence of each faculty on the learning platform.

This visualization is useful for the eLearning committee and the coordinators, to understand that the most important activities of the students are timely information and learning resources distributions. They can drill-down into the data, and understand more how each individual module of each domain fares in the information and resources distribution. This information could be used to provide feedback and guidance to the faculty and students about strategies how to optimize their eLearning processes.

These three use cases complemented with visualizations show that the analytics tool can provide not only overview, but also insight and information about the implementation of the blended learning initiatives in HEI. Additionally, we do not provide possible solutions, actions, or decisions after we have presented the analytics, but leave this step for the users from the stakeholder groups.

5 Next Steps

This is the first iteration in the development cycle of the analytics prototype. We have put considerable thought to meet the expectations and the goals of our stakeholders, but we have not evaluated it with them yet. Thus, the fourth phase of the process, findings confirmations, is missing. However, we have already conducted some unofficial testing and test-usage with different people from the user groups, and the first reactions are very positive. Additionally, we are submitting a demo to the DeLFI conference and we would like to use this opportunity to collect additional feedback about our prototype. Currently, we are planning and preparing the evaluation of the prototype and a pilot phase with selected users. The plan is to develop several iterations of the prototype, and use the pilot phase to test different interfaces, indicators, and visualizations. The goal is to transform the tool from a prototype to production tool/service. The participants in our pilot phase will be blended learning coordinators, developers, and the team for support and qualification. Results from this evaluation phase can be presented at the conference demo and presentation.

6 Conclusion

In this paper, we presented a practical approach, how to introduce data-driven decision making in higher education in Germany. We started by examining the benefits of data-

driven decision making and identifying that this is still a novel approach with development potential. Furthermore, we recognized the different stakeholder groups from the administration that will benefit from tools that can provide actionable intelligence and help them create informed decisions. We collected the perspectives and goals of these user groups, and discovered that they did not have any tools which could systematically support them in their work. Based on this work, we implemented the AiX Analytics prototype. This analytics prototype uses the log data from the learning platform to provide analytics and actionable intelligence to the identified stakeholder groups, and support them through the decision-making processes regarding the eLearning Initiatives and activities at RWTH Aachen University. We already received positive feedback and reception of the prototype, and currently we are planning a pilot phase and evaluation. These activities should provide feedback and results to improve the prototype, and provide this tool as a service for the administration evaluating their blended learning strategy. Additionally, we are going to share our experiences in analytics evaluation and implementation of data privacy compliant solutions in a production environment.

References

- [CDO07] Campbell, J. P.; DeBlois, P. B.; Oblinger, D. G.: Academic Analytics. *Educ. Rev.* 42, S. 40–57, 2007.
- [Ch12] Chatti, M. A.; et al.: A reference model for learning analytics. *Int. J. Technol. Enhanc. Learn.*, 2012.
- [Cr06] Creighton, T.B.: *Schools and Data: The Educator’s Guide for Using Data to Improve Decision Making* (2nd ed.), 2nd Editio. Aufl.: Corwin, 2006.
- [Dy14] Dyckhoff, Anna Lea: *Action Research and Learning Analytics in Higher Education*, 2014.
- [Ho12] Hollingworth, L.; et al.: Data-driven decision making in higher education: One university’s process of revamping the superintendent licensure program. *J. Res. Leadersh. Educ.* 7/1, S. 78–97, 2012.
- [Ko08] Kowalski, Theodore: *Data-driven decisions and school leadership: best practices for school improvement*. Pearson/A and B: Boston, 2008.

A Conversational Agent for the Improvement of Human Reasoning Skills

Laura Wartschinski, Nguyen-Think Le, Niels Pinkwart

Abstract: Human reasoning is the ability to make sound and goal-oriented decisions and is therefore highly relevant in daily life. However, its importance has not yet been addressed explicitly in education. In this paper, we propose to develop a computer-based conversational agent named Liza for improving human reasoning skills. Liza is able to hold conversations with humans to help them solve a small selection of well-studied reasoning problems. Such a conversational agent allows a much more scalable and inexpensive approach for teaching at least basic reasoning principles. The evaluation study shows that the conversational agent Liza improved the reasoning skills of the participants, who had conversations with the agent to solve reasoning problems and that the group using Liza achieved much better learning effects than a group studying with a non-interactive online course that was implemented as a control condition.

Keywords: human reasoning education, pedagogical agent, dialogue system, natural language processing, heuristics and biases

1 Introduction

People are behaving in an instrumentally rational if they pick whatever behavior is best suited to reach their goals, no matter what those goals may be [St12, p.345]. In order to choose the right behavior, it is often necessary to form correct beliefs about the world. Psychological research of the last three decades has shown that people's decisions often deviate from the normatively correct behavior, which is interpreted as irrationality in humans [SW00]. There are many examples for people who behave irrationally, e.g. in displaying errors in calibrating their degrees of belief, not processing information correctly, or failing at assessing probabilities [EO96]. This is mainly because people tend to rely on heuristics that lead to systematic errors, also called cognitive biases [TK74, p. 1124], even when they possess the necessary skills to solve a task correctly [Ka03; K101]. The Dual Process Theory offers an explanation for this behavior [ES13; Ka03]. Furthermore, it is important to note that rationality seems to be a compound quality, as performance on different reasoning tasks shows significant correlation [SW00, p.664], and is also distinct from general intelligence and cognitive ability [SW00; St09]. It has been found that teaching can improve reasoning skills [La04], but in order to do so, requires both repeated explanations and the use of many examples [Ch86, p.293]. Here, the use of educational software may offer benefits. It has been shown that software can successfully be used to train general cognitive skills [Ha11], and educational games [Mo15] as well as videos and learning software [Ge00, p.132] have already been used to improve their user's reasoning. Pedagogical agents, however, who have proven

themselves to be useful as learning motivators and guides [KB07], have not been applied on this subject so far. However, the applicability of many reasoning skills to everyday life predestines the topic for learning in a dialogue-like environment as it is provided by a pedagogical agent. Also, there are not many human teachers available for the field. Therefore, the application of pedagogical agents as teachers for reasoning seems worthwhile. In this work, we prove that a pedagogical agent can improve its users' performance on classical reasoning tasks. Furthermore, we show it has a stronger impact on their performance than a non-interactive online course on the same subject.

2 State of the Art

The goal of this work is to create an Intelligent Tutoring System (ITS) for the teaching of reasoning. ITS are computer programs that interact with students, guide them through the process of learning and provide personalized instructions, tasks and feedback to them. They strive to improve the learning of their students by delivering adaptive and personalized content [VR14]. ITS can be accompanied by conversational agents that engage in chat-oriented or task-oriented dialogues with the user [BL12]. The software system developed in this work also aims to fulfill the role of a Pedagogical Agent, that is, a computer-based agent for educational purposes that delivers information, teaches its users, reacts to its environment (specifically the user input) and chooses the appropriate actions. Current examples for already realized pedagogical agents with similar goals can be found in [Sh09], [ER14] and [La13]. None of those agents however was so far aimed at teaching a general cognitive skills like reasoning. In fact, researchers have only started to examine how modern media could be used in teaching reasoning. For teaching critical thinking, a related composite skill covering argumentation, informal logic and correct judgment, it has been found that computer-assisted training has been superior to training that was carried out traditionally [Hi03, p.188; Ge01, p.547]. Psychologists have already expressed an explicit wish for an interactive software as a teaching tool for this subject [Wa00, p.37].

Previous work on pedagogical agents as tutors for other areas shows very promising result. The agents engage students via social interactions and encourage them to invest effort and persist in learning [KB07], motivating them and keeping them interested [Mo01]. They are able to increase the pleasantness of the learning experience so that user enjoy working with them [Le97]. In general, the use of a pedagogical agent as a social model to improve the motivation and attitude of the student towards the subject is especially effective and makes students face challenges, put effort in their work and persist in learning [KB07]. Students that are taught by agents spend more time with their tasks and easily acknowledge their own mistakes [Ch09]. Those benefits are expected to also work in favor of a pedagogical agent for teaching reasoning and are now applied to this subject for the first time.

3 Design, Content, and Implementation

3.1 Design

This work aims to create a conversational agent, called Liza, which provides natural and, to a certain degree, free communication between computer and user. A text-only approach was chosen to invoke a stronger social presence [DMA05, p.13] and avoid irritations and too high expectations caused by animations that might disappoint the user [Gu11]. Chat-like text pieces and additional images for illustration purposes are therefore the chosen means of communication.

The agent provides explanations on seven topics in the field of reasoning, heuristics and biases, and guides the user through a series of questions, reviewing their level of understanding and correcting their mistakes while at the same time maintained a somewhat natural conversation. It is task-oriented, but uses human social behavior and emotions to improve the learning experience and results. Since effective pedagogical agents suggest correct solutions, provide hints and explanations, give examples, reference to relevant background material and also test the student's abilities, the agent described here implements all those requirements. It aims for a mentor-like role [BK05; Ba00] as a trustworthy and guiding expert that is nevertheless on the same level as the student. Furthermore, the proposed agent uses small talk about personal preferences and sometimes jokes to lighten the mood, as this has been shown to reduce stress [MAM98] and increase enjoyment [BSY09] and a good relationship between agent and user [Gu11; Ku07]. The agent refuses to engage in conflict and stays professional and calm. It does display 'emotions', however. Emotionally responsive agents improve self-efficacy and interest in students [KB07], and the use of emotion makes the interaction much more natural, even when only basic emotions are used [Pi03]. Therefore, Liza shows various sentiments, including pride, joy, curiosity, interest, playfulness, disappointment, sadness and even insecurity, with an emphasis on positive emotions. The agent displays a clear interest in the student's learning success and frequently praises and encourages the users [RN96, p. 310; Sh09]. Praise and flattery expressed by the agent can have a very powerful effect, making the user feel motivated, capable and happy [FN97]. The agent's persona is that of an inexperienced piece of software that tries to learn how to be a good teacher. This narrative allows for a more laid-back style of conversation and gives the agent a plausible goal and also a convenient excuse for any errors.

The agent starts out with a greeting phase in which it explains the process and starts to learn things about the user. Then, during the main part of the program, fourteen questions from the seven different reasoning tasks are presented (two from each). When a task category appears for the first time, an introduction is given, which is accompanied by helpful images and options to give more in-depth explanation. Then the question is asked. For each such question, Liza gives some connecting transition that leads naturally to the topic, then describes a situation and finally asks the question. Knowledge about the user's preference is used to personalize the stories and introductions incorporating the

student's hobbies or personality traits. The user's answer can be decisions, picking several from different options, giving reasons, statements or numbers. If the answer was correct, the agent displays joy and is proud of the user, if not, it might question its own abilities as a teacher and offer to explain the problem again. The user's confidence in their own answers is also determined and later used to provide feedback about their calibration. In between the questions, there are several options for off-task chat. In the end, the agent provides a detailed evaluation of the user's success.

An exemplary piece of dialog out of the middle of the dialogue can look like the following. Liza: *"Are you familiar with the sunk cost fallacy?"* User: *"Never heard of that."* Liza: *"Wonderful! This is my chance for being a good teacher!"* (Liza goes on and explains the concept, then makes sure that the user understood it, then asks the question. The user gives a wrong answer) Liza: *"Hm, I think you might be wrong this time. Do you want me to tell you why?"* User: *"Yeah, go ahead"* (Liza explains the solution, the user agrees) Liza: *"By the way, was that a hard question for you?"* User: *"Probably. I have to wrap my head around it."* Liza: *"Thank you, I noted that down. It's a pleasure to work with you!"* User: *"I like you."* Liza: *"That's so nice! Oh, the next question is about your favority city, Amsterdam. Are you ready?"*

The conversation between user and agent is created as a mixed-initiative dialogue. The agent does not wait forever for a user to answer, but uses flexible waiting times similar to the way a human would act in a conversation. It starts asking worried questions if the user does not react for a very long time, and it offers to give a hint if the answer takes a while. It also reacts to inputs at any time, even if the user just chimed in an insult or a question while the agent was talking. However, Liza does not offer completely free off-task conversation, but rather just answers with short but appropriate messages to whatever was said. Still, those reactions seem sufficiently natural to correspond to the image of a teacher that is largely concentrated on the task. During the whole process, the user can always intervene, for example to stop the dialogue entirely. The direction the conversation takes is changed dynamically to accord for the user's progress and results. The agent picks its replies and questions from a sufficiently large database that stores ten to twenty different variants of reactions for every situation to ensure a natural feeling for the conversation without repetition.

3.2 Content

The content of the agent was taken from the most well-known reasoning tasks for which studies already show that improvement through teaching and exercises is feasible. The seven topics are: Bayesian reasoning [Be81; HG98], the Law of Large Numbers [TK74], the Gambler's Fallacy [K101], Wason's Selection Task [Wa68; K101], Covariance Detection [St09; K101], the Sunk Cost Fallacy [LMN90; Th80] and Belief Bias in Syllogistic Reasoning [EBP83; MN89]. To give one example: The sunk cost fallacy is the irrational desire to stick with the outcomes of previous decision because something was already invested (the 'sunk cost'), even if the continuing commitment provides less

benefit than another option. One example task used by the agent would be started with a short intro about a trip the user takes to a city he likes (inserting the user's favorite city, if known to the agent), and then explaining how he went to a local cinema and paid a lot for the ticket. The agent states that the movie is very boring, that the user is also the only person watching it. Liza describes that the user could spend their time better by doing some other activity. Now it's the user's turn to decide. If the answer is to stay in the cinema because the ticket was costly and is already paid, this is counted as incorrect and an example of the sunk cost fallacy. If the user decides to leave the cinema, this is counted as correct.

3.3 Implementation

Liza was programmed from scratch in Java and was developed using an object-oriented approach. It consists of several units. The Control Unit is responsible for the general management of the dialogue—it keeps track of the state of the conversation, decides what to do next and interacts with all the other components. Input and output are handled by the UI in form of a typical chat interface. The input by the user is then passed on to the Parser, which determines the content or meaning and returns that to the Control Unit, which in turn can decide to react with a phrase from the Phrase Base. The stories are retrieved from the Story Store. A story is a task that is to be presented to the user, complete with descriptions, questions, reactions, explanation and everything else that is necessary for processing. While the Control Unit is executing the stories, correct and incorrect responses are stored by the User Evaluation.

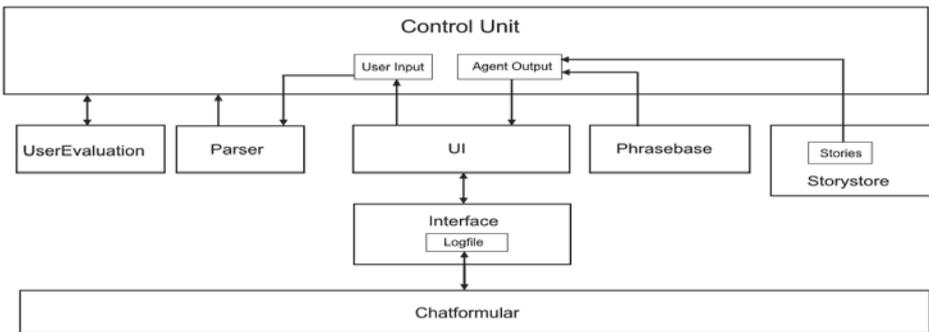


Fig. 1: simplified software architecture

The parsing module relies on basic pattern matching, but is highly specialized for its core purpose: to assess whether the specific questions asked by the agent were answered correctly. Every input is parsed with regard to the context of the message, and more than 80 different context types are distinguished. Knowing the specific context, the agent can be relatively sure that a certain keyword translates to a right or wrong answer to the question asked, whereas without the context, the agent would be clueless because of its rather simple parsing system. This approach reduces the complexity of the parsing a lot

and can be seen as a shortcut to achieve more human-like results with simple methods. The input is parsed recursively with respect to phrases that alter the meaning, like “No” or “wouldn’t”, looking for specific keywords. The combination of keywords that are found and that point to a correct, incorrect or undecided answer determines the final interpretation of the input. Of course, the agent will ask for clarification if the parsing failed. If the user switches the context to something not foreseen by the agent's programming, it will not be able to understand it. For example, if a user asked the agent about the weather, it would not ‘understand’ it, but, depending on the context, glance over it or ask for a rephrasing fitting to the question asked. Partly, this problem is softened by general answers that are interpreted as fitting to the context even if they are not (“don’t you think I should be asking the questions?”). The described parsing mechanism is used for any user input. Since the agents always know the context, it can look out for numbers or verbal statements of quantities or probabilities (“almost certain!”) after asking for the likeliness of a result, and look for affirmation or rejection after asking a yes/no question. That way, the parser can have a very good guess on what the user wanted to communicate.

4 Evaluation

4.1 Study Design

To evaluate the impact of the agent, the following hypotheses were tested: 1. The participants do better on the reasoning tasks after they practiced with the agent than they did before, suggesting they improved their reasoning skills. 2. Talking to the agents leads to a stronger improvement than the non-interactive online course, therefore in the second test, the performance of the participants who talked to the agent will be significantly better than the performance of the participants who took the online course.

The survey for evaluating the success of the agent was conducted using a privately set up server and php and html forms. For each participant, some general information was collected: age, gender, degree of education and level of English skills. The following study design was used: The test subjects were randomly distributed in a treatment group and a control group. The treatment group got to talk to the agent using an online chat interface. The control group read a short text about each of the biases tasks and learned about the underlying principle of reasoning. The texts were taken from different text books and online courses used in actual teaching. For each participant, a test before the intervention and another one afterwards were administered and the performance for every single task was collected. The subjects were forced to pick an answer even if they were not totally sure. This was done to assess even slight preferences to one pick, despite subjects being not totally sure about their reply. For every topic that was assessed with a task, the mean performance before and after the treatment was calculated. For every task, a correct solution was evaluated to have the value 1, an incorrect answer was assigned the value 0. The instructions read as following: “Please solve the following questions by

giving the answers that seem most reasonable, sensible or logical. It's perfectly normal that some of the answer may seem obvious (they probably are—there's no mean trick behind them), and others are harder. Feel free to use a calculator or make notes if it helps you, but please solve the questions on your own. In general, just trust your good reason!" The questions themselves were either the exact same or, if this was not possible, closely modeled after the questions used in current research on reasoning. Participants were obtained by inviting acquaintances and fellow students. 25 more subjects however were paid participants from an online platform called Clickworkers, and 14 were paid participants from various online communities that were paid with \$3 gift cards. In total 65 test subjects completed the survey. The average age of participants was 28.9 years, 44.6% were female, and 82% were fluent or native English speakers. Roughly half of the participants had a Bachelor's degree or a higher education, which is in part due to the acquisition process.

4.2 Results

A two-sample test for proportions was conducted for the treatment group (that was talking to the agent), to test the hypothesis that performance after the treatment would be different to the performance before the treatment. The same test was done for the control group. Also, the change in performance was determined, comparing the mean before the test to the mean afterwards, still assigning a "1" to a correct and a "0" to an incorrect result. For the treatment group, five of the seven tasks showed a significant or very significant improvement, while the control group only improved significantly in one of the tasks. The five significantly ($p < 0.05$) or highly significantly ($p < 0.01$) improved tasks and their improvements were: Sunk Cost Fallacy, Gambler's Fallacy, Bayesian Reasoning, Regression to the Mean and Covariance Detection (see Table 1).

Furthermore, a t-test was conducted comparing the results of the treatment group in the second test (after talking to the agent) to the control group in the second text (after reading their texts), to see if there was any significant difference between them. As it turned out, the treatment group was significantly or very significantly better in the second test than the control group in four of the tasks. There was no task in which the control group performed better in the second task than the treatment group, regardless of significance. A linear regression analysis was performed and showed that none of the demographic variables had any influence on the improvements gained through working with the agent or taking part in the control group's online course. Not even the level of English accounted for any significant change in the effectiveness of the interventions.

Feedback from participants suggested that they perceived the study as interesting. Some gave feedback like "The chatbot is really nice :)" or "Thank you for this interesting experience, I really liked Liza and her funny and helpful way of teaching. :)". Those answers suggest that the people giving them had a positive social experience with the conversational agent. Also, some participants expressed ideas on how to further improve the agent. During the interaction, the users did mainly focus on the dialogue structure the

agent provided and showed no sign of inappropriately high expectations. They stuck to the tasks they were given by Liza and followed her guidance easily.

Task Category	Cond.	Treatment		Control	
		Mean	Std. Err.	Mean	Std. Err.
Sunk Cost Fallacy	before	0.80	0.07	0.63	0.09
	after	1.00	0.00	0.73	0.08
	differ.	0.20**	0.07	0.10	0.12
Gambler's Fallacy	before	0.83	0.06	0.70	0.08
	after	0.97	0.03	0.83	0.07
	differ.	0.14*	0.07	0.13	0.11
Bayesian Reasoning	before	0.17	0.06	0.07	0.05
	after	0.31	0.08	0.23	0.08
	differ.	0.14*	0.07	0.17*	0.09
Belief Bias in Syll. Reasoning	before	0.74	0.07	0.57	0.09
	after	0.63	0.08	0.53	0.09
	differ.	-0.11	0.11	-0.03	0.13
Regression to the Mean	before	0.34	0.08	0.50	0.09
	after	0.77	0.07	0.63	0.09
	differ.	0.43**	0.11	0.13	0.13
Covariation Detection	before	0.71	0.08	0.50	0.09
	after	0.89	0.09	0.53	0.09
	differ.	0.17*	0.11	0.03	0.13
Wason's Selection Task	before	0.40	0.08	0.37	0.09
	after	0.57	0.08	0.30	0.08
	differ.	0.17	0.12	-0.07	0.12

Tab. 1: comparison of treatment (talking to the agent) vs. control (online course) group

4.3 Discussion

This agent was created hoping that the interaction with it would improve the ability of the participants to solve the reasoning tasks. For five tasks, there was a significant increase in performance. The two remaining tasks had no significant effect, although the performance on the selection task did noticeably improve. Therefore, the first hypothesis is largely confirmed. The work with the agent also proved itself to be more efficient than the control group in many cases. The gains in performance were stronger and significant in five cases, whereas the control group could only significantly improve in one task. It can therefore be assumed that the agent provided a more efficient intervention, confirming the second hypothesis.

Only in bayesian reasoning, the control group improved more than the treatment group, but both effect sizes were very close to each other and the task was generally difficult to

teach. Both tasks with the more disappointing, non-significant results (Bayesian Reasoning and Selection Task) are already considered belonging to the hardest reasoning tasks, having generally very low rates in being solved correctly, and being hard to explain, it can be assumed that a short training was not enough to convey the intended message. After all the subjects spent on average 35-45 minutes with the agent, but this time had to be split between the seven different topics they learned about and the general explanation and small talk parts, leaving only roughly 5 minutes on average per topic.

More participants than the acquired 65 would of course offer a more accurate picture of the observed effects. A restriction for the execution of the study was the use of the English language by the bot, which required the participants to be able to write and read in English. The study was conducted online, allowing for participants from different locations to take part in it. It seems highly likely that this allowed for more participants to be acquired, but the sample was not representative. A bias towards younger people is inevitable. Furthermore, many of the participants were motivated extrinsically via payment, in contrast to a real-world application of the agent in a teaching setting without monetary reward. Probably the strongest limitation was the harsh time limit, as none of the participants could be expected to spend more than an hour on the whole process. Therefore, the interaction with the agent was limited in duration, restricting the time for teaching, and it was not possible to assess every task in the test before and after the intervention with a whole set of questions (instead of only one), since that would have further increased the duration of the study. It would have been very interesting to analyze the long-term effects, but only six participants signed up for a follow-up test some weeks later. They all performed very well, but this small sample is not significant in any way.

5 Conclusion and Future Work

The contribution of this paper is two-fold. First, we presented a conversational agent that was designed to improve skills for a selection of core reasoning tasks and to the best of our knowledge, this learning domain has not been addressed in the community of technology-enhanced learning. The second contribution is the empirical study of the developed conversational agent for human reasoning. Although positive impact could not be found in solving all the tasks, maybe because Wason's Selection Task and Bayesian reasoning are both very complicated concepts to teach, the evaluation study showed very satisfying results with most of the tasks. The agent demonstrated to be clearly more effective than participating an online course. It can be concluded that performance on solving certain reasoning tasks can indeed be improved with the help of pedagogical agents, which is especially fortunate as there are not many teachers in this area, despite its importance for arguably everybody's personal life.

Given the nature of the agent, it can always be improved by adding more special parsing options and appropriate responses. By extending the scope of correctly parsed input, the

agent can become even more convincing, although there might be diminishing returns for increasingly rare topics. More elaborate dialogue patterns could be added, too. Further improvements could be made by allowing the user to ask the agent for more questions of a certain kind, or the agent suggesting more training in an area where the user seems to lack understanding. The agent could probably be put to a more effective use if it was used not only for a single intervention, but regularly. Storing the past performances of users, remembering their preferences, and specifically targeting the areas where they can improve come to mind as obvious possibilities for expanding the agent. Also, right now the agent does only teach students in seven limited and strictly restricted topics. Improvements could include more topics (e.g. addressing some more of the wide variety of known fallacies and biases), offering a more general approach (e.g. giving some philosophical background on knowledge, its representation and a more refined approach on what rationality means), or digging into new areas of the field (e.g. the area of ‘critical thinking’ which includes open-mindedness and discussing the quality of arguments and which was originally intended to be part of the agent, but had to be left out due to time and scale concerns). By those means, the ‘curriculum’ of the agent could be extended in many directions. First and foremost, it would be very interesting to validate the results with a much larger sample than the 65 participants who were available. An expansion of this research could be a study over the course of several weeks or months, to have the users spend much more time with the agent, to see if this approach is more effective. Long term effects could be investigated by reassessing user’s performance after weeks, months or even years.

References

- [Ba00] Baylor, A.: Beyond butlers: Intelligent agents as mentors. *Journal of Educational Computing Research* 22/4, S. 373–382, 2000.
- [Be81] Berwick, D.M. et al.: When doctors meet numbers”. *The American journal of medicine* 71/6, S. 991–998, 1981.
- [BK05] Baylor, A.; Kim, Y.: Simulating instructional roles through pedagogical agents. *Int. Journal of Artificial Intelligence in Education* 15/2, S. 95–115, 2005.
- [BL12] Banchs, R.E.; Li, H.: IRIS: a chat-oriented dialogue system based on the vector space model. In: *Proceedings of the ACL System Demonstrations*, 2012.
- [BSY09] Bickmore, T.; Schulman, D.; Yin, L.: Engagement vs. deceit: Virtual humans with human autobiographies”. In: *Int. Workshop on Intelligent Virtual Agents*. S. 6–19, 2009.
- [Ch09] Chase, C.C. et al: Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of Science Education and Technology* 18/4, 2009.
- [Ch86] Cheng, P.W. et al.: Pragmatic versus Syntactic Approaches to Training Deductive Reasoning. *Cognitive Psychology* 18, S. 293–328, 1986
- [DMA05] Dirkin, K.H.; Mishra, P.; Altermatt, E.: All or nothing: Levels of sociability of a

pedagogical software agent and its impact on student perceptions and learning. *Journal of E. Multimedia and Hypermedia* 14/2, 2005.

- [EBP83] Evans, J.; Barston, J.L.; Pollard, P.: On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition* 11/3, S. 295–306, 1983.
- [EO96] Evans, J.; Over, D.E.: *Rationality and Reasoning*. Psychology Press, 1996.
- [ER14] Epstein, D.; Reategui, E.: A pedagogical agent with embedded data mining functions to support collaborative writing. *Learning* 12/13, S. 14, 2014.
- [ES13] Evans, J.; Stanovich, K.E.: Dual-process theories of higher cognition advancing the debate”. *Perspectives on psychological science* 8/3, 2013.
- [FN97] Fogg, B.J.; Nass, C.: Silicon sycophants: the effects of computers that flatter. *Int. Journal of Human-Computer Studies* 46/5, S. 551–561, 1997.
- [Ge00] Gelder, T. van: Learning to reason: A Reason!Able approach”. In: *Proc. of the Fifth Australasian Cognitive Science Society Conference*. Adelaide: Causal, 2000.
- [Ge01] Gelder, T. van: How to improve critical thinking using educational technology. In: *Proc. of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, S. 539–548, 2001.
- [Gu11] Gulz, A. et al.: Building a Social Conversational Pedagogical Agent. In: *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*, S. 128, 2011.
- [Ha11] Haferkamp, N. et al.: Training disaster communication by means of serious games in virtual environments. *Entertainment Computing* 2/2, S. 81–88, 2011.
- [HG98] Hoffrage, U.; Gigerenzer, G.: Using natural frequencies to improve diagnostic inferences. *Academic medicine* 73/5, S. 538–540, 1998.
- [Hi03] Hitchcock, D.: The effectiveness of computer-assisted instruction in critical thinking. *Informal Logic: Reasoning and Argumentation in Theory and Practice* 24/3, S. 183–217, 2003.
- [Ka03] Kahneman, D.: A perspective on judgment and choice: mapping bounded rationality. *American psychologist* 58/9, 2003.
- [KB07] Kim, Y.; Baylor, A.: Pedagogical agents as social models to influence learner attitudes. *Educational Technology* 47/1, S. 23–28, 2007.
- [KI01] Klaczynski, P.A.: Analytic and Heuristic Processing Influences on Adolescent Reasoning and Decision-Making. *Child development* 72/3, S. 844–861, 2001
- [Ku07] Kumar, R. et al.: Supporting students working together on math with social dialogue. In: *Workshop on Speech and Language Technology in Education*, S. 96–99, 2007.
- [La13] Lane, H.C. et al.: The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In: *Int. Conference on Artificial Intelligence in Education*, S. 309–318, 2013.
- [La04] Larrick, R.P.: Debiasing. In: *Blackwell handbook of judgment and decision making*, S. 316–337, 2004.

- [Le97] Lester, J.C. et al.: The persona effect: affective impact of animated pedagogical agents. In: Proc. of the ACM SIGCHI Conference on Human factors in computing systems, S. 359–366, 1997.
- [LMN90] Larrick, R.P.; Morgan, J.N.; Nisbett, R.E.: Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science* 1/6, S. 362–370, 1990.
- [MAM98] Mulken, S.van; André, E.; Müller, J.: The persona effect: how substantial is it?” In: *People and computers XIII*. Springer, S. 53–66, 1998.
- [MN89] Markovits, H.; Nantel, G.: The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition* 17/1, S. 11–17, 1989.
- [Mo01] Moreno, R. et al.: The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction* 19/2, S. 177–213, 2001.
- [Mo15] Morewedge, C.K. et al.: Debiasing Decisions Improved Decision Making With a Single Training Intervention. In: *Policy Insights from the Behavioral and Brain Sciences* 2/1, S. 129–140, 2015.
- [Pi03] Picard, R.W.: Affective computing: challenges. *Int. Journal of Human-Computer Studies* 59/1, S. 55–64, 2003.
- [RN96] Reeves, B.; Nass, C.: *How people treat computers, television, and new media like real people and places*. CSLI Pub. and Cambridge University Press, 1996.
- [Sh09] Shen, E.: Effects of agent emotional support and cognitive motivational messages on math anxiety, learning, and motivation, 2009.
- [St09] Stanovich, K.E.: *What intelligence tests miss: The psychology of rational thought*. Yale University Press, 2009.
- [St12] Stanovich, K.E.: On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In: *The Oxford handbook of thinking and reasoning*, S. 343–365, 2012.
- [SW00] Stanovich, K.E.; West, R.F.: Individual differences in reasoning: Implications for the rationality debate? *Behavioral and brain sciences* 23/5, 2000.
- [Th80] Thaler, R.: Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization* 1/1, S. 39–60, 1980.
- [TK74] Tversky, A.; Kahneman, D.: Judgment under uncertainty: Heuristics and biases. *Science* 185/4157, S. 1124–1131, 1974.
- [VR14] Veletsianos, G.; Russell, G.S.: *Pedagogical agents*. In: *Handbook of research on educational communications and technology*. Springer, S. 759–769, 2014.
- [Wa00] Walton, D.: Problems and useful techniques: My experiences in teaching courses in argumentation, informal logic and critical thinking. *Informal Logic* 20/2, 2000.
- [Wa68] Wason, P.C.: Reasoning about a rule. *The Quarterly journal of experimental psychology* 20/3, S. 273–281, 1968.

Ein Lückentext-Test zur Beherrschung einer Programmiersprache

Michael Striewe¹, Matthias Kramer² und Michael Goedicke³

Abstract: Dieser Beitrag präsentiert und diskutiert erste Erfahrungen mit einem experimentellen Ansatz, der das sprachwissenschaftliche Konzept der C-Tests auf das Verständnis von Programmiersprachen überträgt. Es werden sowohl die Konstruktion der Aufgaben als auch erste vorläufige Ergebnisse erläutert, die eine weitere Untersuchung dieses Aufgabentyps motivieren.

Keywords: Programmierausbildung, Codeverständnis, C-Test, Lückentext, Kompetenzmessung

1 Einleitung

Programmieranfänger sehen sich beim Erlernen ihrer ersten Programmiersprache verschiedenen Herausforderungen ausgesetzt: Das allgemeine Paradigma der gewählten Sprache muss verstanden werden, Programmverhalten muss in Datenstrukturen, Datenflüsse und Kontrollflüsse innerhalb dieses Paradigmas umgesetzt werden und nicht zuletzt muss die Syntax der Sprache beherrscht werden. Letztgenannter Punkt gilt genauso für erfahrene Programmierer, die eine neue Programmiersprache lernen, selbst wenn sie dabei an einem vertrauten Programmierparadigma festhalten.

Dass das Erlernen der Syntax eine Barriere für Anfänger sein kann, ist gut bekannt [Mc01, SS13]. Daher ist es teilweise möglich, aus der Art und Häufigkeit von Syntaxfehlern bei der Programmkonstruktion auf das Niveau der Sprachbeherrschung zu schließen [SH15]. Für einen diagnostischen Test erscheint dieses Vorgehen aber zu ungenau, da die Programmkonstruktion im Rahmen einer Programmieraufgabe neben der reinen Beherrschung der Programmiersprache auch andere Anforderungen stellt. Es erscheint daher wünschenswert, ein Testinstrument zu entwickeln, mit dem Schwächen in der Syntaxbeherrschung ermittelt werden können, ohne dass die Lernenden dazu eigenständig vollständigen Programmcode erzeugen müssen. Entsprechende Aufgaben könnten dann als Ergänzung zu klassischen Programmieraufgaben eingesetzt werden.

Aus den Sprachwissenschaften sind für diesen Einsatzzweck sogenannte C-Tests bekannt, um das Niveau der Beherrschung einer (Fremd-)Sprache zu bestimmen [KR82]. Sie werden z.B. eingesetzt, um Anfängern einen Sprachkurs auf dem richtigen Level

¹ Paluno - The Ruhr Institute for Software Technology, Universität Duisburg-Essen, Campus Essen, michael.striewe@paluno.uni-due.de

² Didaktik der Informatik, Universität Duisburg-Essen, Campus Essen, matthias.kramer@uni-due.de

³ Paluno - The Ruhr Institute for Software Technology, Universität Duisburg-Essen, Campus Essen, michael.goedicke@paluno.uni-due.de

zuzuweisen [K185]. Der Einsatz ist nicht auf eine bestimmte Sprache oder Sprachfamilie beschränkt, sondern wurde in zahlreichen Sprachen mit verschiedenen Notationen verwendet [Ba14]. Auch wenn Programmiersprachen formale und keine natürlichen Sprachen sind, erscheint es aufgrund der positiven Erfahrungen mit C-Tests bzgl. Reliabilität, Validität und Objektivität und trotz der existierenden Kritik an diesen Tests [BM15] vielversprechend, das Konzept auch hier zu erproben. Der vorliegende Beitrag demonstriert daher eine Adaption und experimentelle Erprobung des Konzepts von C-Tests für das Erlernen einer Programmiersprache. Die vorläufigen Ergebnisse zeigen, dass dieser Aufgabentyp tatsächlich diagnostisches Potenzial hat und motivieren so eine weitere Erforschung.

Der Beitrag ist wie folgt gegliedert: Abschnitt 2 diskutiert den Forschungskontext und verwandte Arbeiten. Abschnitt 3 erläutert die Aufgabenkonstruktion für die adaptierten C-Tests für Programmiersprachen. Abschnitt 4 stellt eine Beispielaufgabe und die darauf basierenden bisherigen Evaluationsergebnisse vor. Abschnitt 5 schließt den Beitrag mit einem Fazit und einem Ausblick auf nachfolgende Forschungsaktivitäten ab.

2 Forschungskontext und Related Work

Bereits 1986 beschrieb du Boulay [Du86] die Aspekte, die beim Erlernen von Programmierung parallel eingeübt werden müssen und daher nicht selten in Überforderung und Frustration münden. Neben dem Verstehen des Programmablaufs sowie dem Entwickeln allgemeiner Problemlösungsstrategien spielt das Erlernen einer *Notation*, also der Syntax und Semantik einer Programmiersprache, eine wichtige Rolle. Zum einen stellt der Umgang mit der Syntax nachweislich ein wesentliches Hindernis beim Erlernen einer Programmiersprache dar (siehe z.B. [Mc01] und [SS13]). Zum anderen scheint das Wissen um syntaktische Bausteine und deren Semantik allein nicht auszureichen, um diese zu funktionierenden Programmen zusammensetzen [Wi96]. Der korrekte Umgang mit einer Programmiersprache, d.h. Programmierbausteine syntaktisch und semantisch sinnvoll miteinander kombinieren zu können, scheint daher mutmaßlich eine notwendige Teilfertigkeit zu sein, die für das erfolgreiche Erlernen von Programmierung von wesentlicher Bedeutung ist. Hier wird auch die inhaltliche Nähe zum linguistischen C-Test offensichtlich.

Betrachtet man nun Programmierkompetenz als multidimensionales latentes Konstrukt [KHB16], so erscheint es sinnvoll, durch vielfältige Testitems im Rahmen eines formativen Assessments festzustellen, welche Ausprägungen in den verschiedenen Teildimensionen zu (Miss-)Erfolgen führen und welche daraus resultierenden individuellen Lernbedürfnisse die Studierenden haben. Insbesondere können so bei einem späteren summativen Assessment (bspw. der Abschlussklausur eines Semesters) bessere und gezieltere Rückschlüsse auf generelle Fehlvorstellungen und Probleme gezogen werden, was wiederum zu einer möglichen Verbesserung der Lehre beiträgt.

Es ist klar, dass für eine umfassende Testung von Programmierkompetenz eine bloße

Adaption des C-Tests für die Programmierung allein nicht ausreicht [KI04]. Dennoch kann ein solcher Test als Baustein in einem größeren Assessment wertvolle Hinweise auf die Leistung in Kompetenzfacetten wie *Verstehen gegebener Quelltexte* bzw. *syntaktisches Verständnis einer Programmiersprache* liefern. Zudem stellt er durch das Format eine ideale Schnittstelle zwischen den Fertigkeiten *Quelltexte interpretieren* und *Quelltexte produzieren* dar, was ebenfalls vermuten lässt, dass durch einen solchen Test verschiedene Fertigkeiten abgeprüft werden. Hartig & Harsch konnten für C-Tests in Fremdsprachen bereits empirisch nachweisen, dass multidimensionale Modelle die Varianzen besser erklären können als unidimensionale [HH09, S. 61].

3 Aufgabenkonstruktion

Das Konzept der C-Tests in ihrer üblichen Form sieht vor, in einem Text beginnend mit dem zweiten Satz in jedem zweiten Wort die hintere Hälfte des Wortes durch eine Lücke zu ersetzen. Das Füllen einer solchen Lücke mit einer korrekten Lösung erfordert dabei sowohl Kenntnisse im Vokabular der Sprache, um das fragmentierte Wort zu erkennen oder aus dem Kontext zu erschließen, als auch in vielen Fällen der Grammatik, um die passende Endung zu bestimmen. Eine direkte Übernahme dieses Konzeptes erscheint für Programmiersprachen nicht sinnvoll, da das Vokabular eines Programms in der Regel so aufgebaut ist, dass sich aus dem ersten Teil eines Wortes sofort ohne Kenntnis von Grammatik und Kontext der fehlende Teil ergibt (z.B. bei Schlüsselworten), oder aber der zweite Teil nahezu willkürlich gewählt werden kann (z.B. bei Bezeichnern). Ferner treten im Programmcode nicht nur Worte im klassischen Sinne auf, sondern auch Operatoren für Zuweisungen, Vergleiche oder logische Verknüpfungen, die eine wichtigere Rolle haben als Satzzeichen in einem natürlichsprachlichen Text.

Daher wurde eine Adaption des Verfahrens gewählt, bei der eine gewisse Menge von Wörtern komplett in eine Lücke umgewandelt wird. Als Wörter werden dabei Schlüsselwörter sowie weitere beliebige Zeichenketten behandelt, die spezifisch pro Programmiersprache oder sogar pro Quelltext zu definieren sind. Im Kontext von Java erscheint es beispielsweise sinnvoll, Zeichenketten wie `String`, `System` oder `println` als ersetzbare Worte zu behandeln, auch wenn sie keine Schlüsselwörter der Programmiersprache, sondern vorgegebene Typen und Methoden der Java API sind. Ferner kann es sinnvoll sein, Operatoren wie `==` oder `&&` sowie beliebige Klassennamen in Lücken umzuwandeln, sofern deren Lösung aus dem Programmcode geschlossen werden kann.

4 Beispielaufgabe und Evaluation

Eine auf diesem Ansatz basierende Beispielaufgabe zur Programmiersprache Java mit insgesamt 13 Lücken unterschiedlichen Charakters zeigt Abbildung 1. Sie zeigt einen Screenshot aus dem zur Evaluation verwendeten E-Assessment-System in der Ansicht für Lernende.

Betrachten Sie den folgenden Programmcode, der an einigen Stellen Lücken enthält. Ergänzen Sie diese Lücken sinnvoll.

```

public class Bank {
    public static final int MAX_KONTEN = 10;
    int anzahl = 0;

    Konto[] konten = [ ] Konto[MAX_KONTEN];

    private Konto kontoSuchen(int kontoNr) {
        for (int i = 0; i < konten.length; i++) {
            [ ] (konten[i]!=null && konten[i].getKontoNr() == kontoNr) {
                [ ] konten[i];
            }
        }
        return null;
    }

    public boolean kontoAnlegen(Konto neuesKonto) {
        if (kontoSuchen(neuesKonto.getKontoNr()) == null && anzahl < MAX_KONTEN) {
            for (int i = 0; i < konten.length; i++) {
                if (konten[i] [ ] null) {
                    konten[i] = neuesKonto;
                    anzahl++;
                    return true;
                }
            }
            [ ] false;
        }
    }

    public boolean kontoLoeschen( [ ] kontoNr) {
        for (int i = 0; i [ ] konten.length; i++) {
            [ ] (konten[i]!=null && konten[i].getKontoNr() == kontoNr) {
                konten[i] = null;
                anzahl--;
                return true;
            }
        }
        [ ] false;
    }

    public double berechneVermoeigen(Konto[] konten) {
        double vermoeigen = 0.0;
        [ ] (int i = 0; i < konten.length; i++) {
            Konto k = konten[i];
            vermoeigen += k.kontoStand;
        }
        [ ] vermoeigen;
    }

    private [ ] ausgabeAllerKonten() {
        for (Konto i : konten)
            [ ] (i != null)
                System.out.println("Konto " + i.getKontoNr()
                    + " hat einen Kontostand von: " + i.getKontoStand());
    }
}

```

Abb. 1: Screenshot der zur Evaluation verwendeten Beispielaufgabe. Die erwarteten korrekten Antworten sind in der Kopfzeile von Tabelle 1 ersichtlich.

In 11 der Lücken fehlen Schlüsselworte der Sprache Java, in den verbleibenden zwei fehlen Vergleichsoperatoren. Die fehlenden Schlüsselworte sind viermal das `return`-Statement, dreimal `if` und je einmal `void`, `new`, `int` und `for`. Die Lücken erfordern eine unterschiedlich umfangreiche Betrachtung des Programmcodes: Der Einsatz von `return`, `if`, `for` und `new` lässt sich aus der jeweiligen Zeile direkt ableiten, da andere denkbare Einträge in der jeweiligen Lücke keine syntaktisch korrekte Lösung bilden

würden. Für die Lösung der Lücke für `void` ist es dagegen notwendig, die betroffene Methode zumindest soweit zu verstehen, dass erkannt wird, dass in dieser Methode keine Rückgabe erfolgt. Die Lücke für `int` kann schließlich nur nach genauerer Betrachtung des Datenflusses geschlossen werden. Dies erfordert auch Sorgfalt beim Erstellen der Aufgabe, da Lücken mit Datentypen ggf. nicht eindeutig lösbar sind, wenn in einer Methode eine lokale Variable deklariert und anschließend zurückgegeben wird. Werden nun sowohl bei der Deklaration als auch in der Signatur die Typnamen zu einer Lücke geändert, könnte jede Lösung korrekt sein, die beide Lücken identisch füllt. Ebenfalls zur Sicherstellung der Eindeutigkeit wurde beim Erstellen der Beispielaufgabe auf die Tilgung von Sichtbarkeits-Modifiern verzichtet, da die korrekte Lösung dazu ohne weitere Kenntnis des Programmkontextes nicht zu bestimmen ist.

Zur experimentellen Evaluation haben 11 Personen die Beispielaufgabe bearbeitet. Bei diesen handelt es sich um Studienanfänger, Auszubildende und Schüler, die alle geringe bis mäßige Erfahrung (max. ein Schuljahr/eine Vorlesung) in Java hatten. In den Ergebnissen in Tabelle 1 ist ersichtlich, dass sowohl in der Summe als auch in der Verteilung der Fehler unterschiedliche Leistungen erzielt wurden. Beispielsweise wurde Lücke 6 dreimal falsch geschlossen, während bei den vier `return`-Lücken insgesamt nur zwei Fehler auftraten. Ferner gibt es Hinweise auf unterschiedliche Personenparameter, da Testperson 11 alle `if`-Lücken falsch geschlossen hat und alle anderen korrekt, während alle anderen Personen zusammen nur einen Fehler bei einer `if`-Lücke machten.

Person	L1 new	L2 if	L3 return	L4 ==	L5 return	L6 int	L7 <	L8 if	L9 return	L10 for	L11 return	L12 void	L13 if	Σ
#1	1	1	1	1	1	1	1	1	1	1	1	1	1	13
#2	1	1	1	1	1	1	1	1	1	1	1	1	1	13
#3	1	1	1	1	1	0	1	1	1	1	1	1	1	12
#4	0	1	1	1	0	1	0	1	1	1	1	0	0	8
#5	1	1	1	1	1	0	1	1	1	1	1	1	1	12
#6	1	1	1	0	1	0	1	1	1	1	1	1	1	11
#7	1	1	1	1	1	1	1	1	1	1	1	0	1	12
#8	1	1	1	1	1	1	1	1	1	1	1	1	1	13
#9	1	1	0	0	1	1	1	1	1	1	1	1	1	11
#10	1	1	1	1	1	1	1	1	1	0	1	1	1	12
#11	1	0	1	1	1	1	1	0	1	1	1	1	0	10
Σ	10	10	10	9	10	8	10	10	11	10	11	9	9	

Tab. 2: Ergebnisse des Experiments. Eine 1 steht für eine korrekt gefüllte Lücke, eine 0 für eine falsch oder gar nicht ausgefüllte Lücke.

Die geringe Anzahl an Daten verbietet eine weitergehende statistische Analyse. Die Ergebnisse lassen jedoch erwarten, dass weitere Experimente relevante Beziehungen zwischen Personenparametern und Charakteristika der Lücken aufdecken könnten.

5 Fazit und Ausblick

Es konnte gezeigt werden, dass die Konstruktion von C-Tests mit einigen Anpassungen

auch für Programmiersprachen sinnvoll möglich ist. Die vorläufigen Ergebnisse legen nahe, dass das Potenzial der Aufgaben sowie die Einflussfaktoren für die Schwierigkeit der Lösung solcher Aufgaben weiter untersucht werden sollte. Insbesondere ist zu prüfen, ob die Zahl der Fehler ohne weitere Betrachtung der falsch gefüllten Lücken ein gutes Maß für die Beherrschung einer Programmiersprache ist und welcher Quelltext-Umfang mit welcher Frequenz von Lücken als Testwerkzeug geeignet ist. Es ist zu erwarten, dass sich dabei auch Unterschiede je nach Programmiersprache ergeben.

Literaturverzeichnis

- [Ba14] Baghaei, Purya: Construction and validation of a C-Test in Persian. In (Grotjahn, Rüdiger, Hrsg.): *Der C-Test: Aktuelle Tendenzen*, S. 299–312, 2014.
- [BM15] Baur, Rupprecht; Mashkovskaja, Anna: C-Test-Kritik reviewed. In (Böcker, Jessica; Stauch Anette, Hrsg.): *Konzepte aus der Sprachlehrforschung – Impulse für die Praxis*, S. 435-449, 2015.
- [Du86] Du Boulay, Benedict: Some difficulties of learning to program. *Journal of Educational Computing Research*, 2(1), S. 57–73, 1986.
- [HH09] Hartig, Johannes; Höhler, Jana: Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35(2), S. 57–63, 2009.
- [Kl85] Klein-Braley, Christine: C-Tests as placement tests for German university students of English. *AKS-Rundbrief*, 13/14, S. 96–100, 1985.
- [KR82] Klein-Braley, Christine; Raatz, Ulrich: *Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung*. *AKS-Rundbrief* 4, S. 23–37, 1982.
- [KHB16] Kramer, Matthias; Hubwieser, Peter; Brinda, Torsten: A Competency Structure Model of Object-Oriented Programming. In: *Learning and Teaching in Computing and Engineering (LaTICE)*, 2016 International Conference on IEEE, S. 1–8, 2016.
- [Mc01] McCracken, Michael; Almstrum, Vicki; Diaz, Danny; Guzdial, Mark; Hagan, Dianne; Kolikant, Yifat Ben-David; Laxer, Cary; Thomas, Lynda; Utting, Ian; Wilusz, Tadeusz: A multi-national, multi-institutional study of assessment of programming skills of first-year CS students. *ACM SIGCSE Bulletin*, 33(4), S. 125–180, 2001.
- [SH15] Schulten, Bastian; Höppner, Frank: Zur Einschätzung von Programmierfähigkeiten – Jedem Programmieranfänger über die Schultern schauen. In: *Proceedings des Zweiten Workshops „Automatische Bewertung von Programmieraufgaben“*, Wolfenbüttel. 2015.
- [SS13] Stefik, Andreas; Siebert, Susanna: An empirical investigation into programming language syntax. *ACM Transactions on Computing Education (TOCE)*, 13(4), S. 19, 2013.
- [Wi96] Winslow, Leon E.: Programming pedagogy - a psychological overview. *ACM Sigcse Bulletin*, 28(3), S. 17–22, 1996.

SensoMot – Sensorische Erfassung von Motivationsindikatoren zur Steuerung adaptiver Lerninhalte

Oliver Schneider¹, Thomas Martens², Mathias Bauer³, Alexandra Ott-Kroner⁴,
Uwe Dick⁵ und Michel Dorochevsky⁶

Abstract: Im Projekt SensoMot wird untersucht, wie Motivationslagen durch unaufdringliche Sensoren erfasst und die hieraus abgeleiteten Parameter genutzt werden können, um Lerninhalte zur Laufzeit anzupassen. Eine so erhöhte Lernmotivation könnte in technologiebasierten Lehr-Lern-Szenarios zu größeren Lernerfolgen führen.

Keywords: eLearning, adaptives Lernen, Motivation, Sensoren, mobile Computing, Wearables

1 Adaptation für motiviertes Lernen

Motiviertes Lernen ist die Voraussetzung für eine tiefe und vernetzte Verarbeitung von Lerninhalten sowie einer langen Behaltensleistung. Umgekehrt können Störungen der Motivation zu oberflächlichen Lernprozessen oder sogar zu Lernblockaden führen. Eine frühe und vor allem nicht-reaktive Erfassung der Motivationslagen könnte deshalb eine große Hilfe für die Personalisierung von Lernsituationen durch passende Inhaltsadaptation sein. Dies ist die Zielvorstellung von SensoMot: Es soll ermöglicht werden, motivationale Parameter für die Adaptation von Lerninhalten systematisch zu berücksichtigen und damit in technologiebasierten Lehr-Lern-Szenarios die Lernmotivation und die Lernprozesse entscheidend zu verbessern.

Um die Adaption von Lerninhalten mit Hilfe von Motivationsindikatoren erreichen zu können, ist die Identifikation von entsprechenden Motivationslagen nötig. Dies könnte mit Hilfe eines mobilen Endgerätes durch Selbstberichtfragebögen mit der sogenannte „Experience Sampling Method“ [HSC07] erfolgen. In einer Vorstudie wurden die fünf Indikatoren Angespanntheit, Energie, Interesse, Konzentration und Affekt erhoben.

¹ DIPF – Deutsches Institut für Internationale Pädagogische Forschung, TBA – Zentrum für technologiebasiertes Assessment, Schloßstraße 29, 60486 Frankfurt am Main, Oliver.Schneider@dipf.de

² MSH Medical School Hamburg, Fakultät Gesundheit, Am Kaiserkaai 1, 20457 Hamburg, thomas.martens@medicalschoool-hamburg.de

³ Technische Universität Ilmenau, Fachgebiet Medienproduktion, Ehrenbergstraße 29, 98693 Ilmenau, Mathias.Bauer@tu-ilmenau.de

⁴ Fernlehrinstitut Dr. Robert Eckert GmbH, Leitung Produktentwicklung, Dr.-Robert-Eckert-Straße 3, 93128 Regenstauf, Alexandra.Ott-Kroner@eckert-schulen.de

⁵ Leuphana Universität Lüneburg, Institut für elektronische Geschäftsprozesse (IEG), Scharnhorststraße 1, 21335 Lüneburg, Uwe.Dick@leuphana.de

⁶ Nagarro GmbH, ADS/CBA, Aidenbachstraße 42, 81379 München, Michel.Dorochevsky@nagarro.com

Dieses Vorgehen hat mehrere Probleme: diese Selbstauskunft unterbricht den normalen Lernfluss und die Beantwortung der Fragen erfordert Aufmerksamkeit sowie Verarbeitungskapazität, die möglicherweise den eigentlichen Lernprozess stört. Deshalb sollen Selbstberichtsdaten durch nicht-reaktive Sensordaten abgelöst werden, die automatisch verarbeitet und weitergegeben werden. Das Ziel der Forschung in SensoMot besteht in der Bestimmung und Erfassung von motivationalen Parametern durch die Verwendung von nicht-reaktiven Sensordaten für die automatische Steuerung und Adaptation von Lerninhalten. Zur Verarbeitung der Messdaten werden Methoden des maschinellen Lernens eingesetzt. Eine wissenschaftliche Herausforderung besteht in SensoMot darin, dass die genauen Zusammenhänge zwischen den Sensordaten und den gewünschten Zielgrößen (Motivationsindikatoren) noch identifiziert werden müssen.

2 Überblick

Das didaktische Gesamtdesign folgt dem integrierten Lern- und Handlungsmodell (ILHM) [Ma12] und soll vor allem die motivationale Selbstregulation des Lernens unterstützen. Die Passung zwischen Lernenden und Lernsituation soll durch technologische Adaption systematisch verbessert werden. Während in natürlichen Lehr-Lern-Settings die Passung leicht durch eine entsprechend geschulte Lehrkraft verbessert werden kann, ist dies in technologischen Lehr-Lern-Situationen – zumindest bisher – kaum möglich. Ein mit SensoMot-Technologien ausgestattetes System soll in der Lage sein, auf kritische Lernereignisse zu reagieren und die Lernangebote dann zu adaptieren.

Gemäß dem ILHM können grob drei Formen der Adaption unterschieden werden: inhaltliche Adaption durch alternative Inhaltssegmente, methodische Adaption durch alternative Lernzugänge und Adaption der aktuellen Lernvorgänge z. B. durch das Lerntempo. Das Lernsystem sollte automatisch erkennen können, wenn etwa das Interesse (Motivationsindikator) an den Lerninhalten absinkt und könnte ein interessanteres Inhaltssegment (Inhaltsadaption) vorschlagen. Es sollte Anspannung (Motivationsindikator) identifizieren und andere Lernzugänge empfehlen (Adaption der Lernzugänge) oder sollte einen Konzentrationsabfall (Motivationsindikator) feststellen können, der auf einer zu leichten Aufgabenstellung beruht und entsprechend die Schwierigkeit der der aktuellen Aufgaben erhöhen (Adaption aktueller Lernvorgänge).

Solche Adaptionen müssen von den Autoren der Lernsegmente antizipiert und dann mit einer Autorensoftware vorbereitet werden. Dies kann bedeuten, dass das gleiche Inhaltssegment in unterschiedliche Kontexte eingebettet wird und die entsprechende Adaption dann über geeignete Zuordnungsalgorithmen realisiert wird. Als Basis dient der CBA-ItemBuilder [Rö12], der zum einen die grafische Erstellung von adaptiven Lerninhalten ermöglicht (Autorensystem) sowie zum anderen deren Ausführung (Laufzeitsystem).

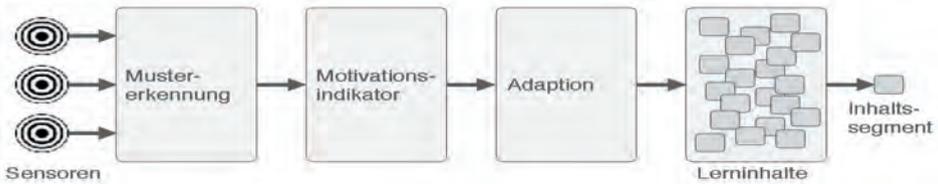


Abb. 1: SensoMot Prozesskette: Sensordaten werden mittels Mustererkennung zu einem Motivationsindikator verdichtet, der für die Adaption von Inhaltssegmenten verwendet wird.

Die Sensordaten müssen erhoben, dann zusammengeführt und interpretiert werden, bevor sie im CBA-ItemBuilder zur Steuerung der Lerninhalte eingesetzt werden können. Die vollständige Prozesskette zeigt Abb. 1. Vielversprechend sind Messungen, die unkompliziert auf der Haut stattfinden können sowie Benutzerinteraktionen mit dem Lernsystem. Abb. 2 zeigt den technologischen Ansatz des Projekts.

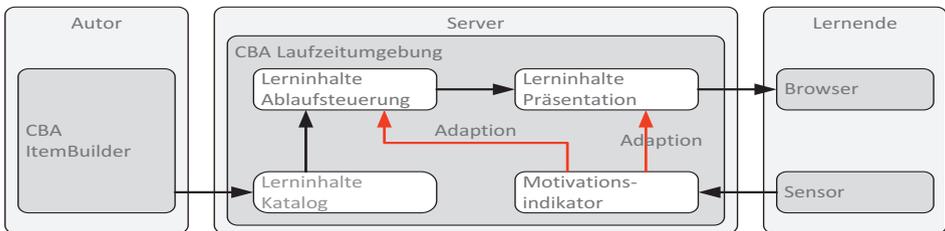


Abb. 2.: Technologischer Ansatz: Der mittels Sensoren gewonnene Motivationsindikator steuert die Adaption der Lerninhalte, welche im CBA-ItemBuilder erstellt werden.

Das Projekt SensoMot folgt einem nutzerzentrierten Entwicklungsansatz, bei dem die späteren Benutzer kontinuierlich in den Entwicklungsprozess einbezogen werden [HP12]. Dazu müssen Motivationsdaten auf eine für den Benutzer einfache Weise gesammelt werden. Die Evaluation erfolgt projektbegleitend in mehreren Zyklen auf Grundlage des Design Based Research-Ansatzes [Re05]. Die Ergebnisse werden in die Steuerung des Projekts einfließen. Die Qualitätssicherung bezieht sich auf die systematische Überprüfung der sensorischen Mustererkennung, die Datentriangulation und die motivationale Adaption der Lerninhalte sowie die Nutzbarkeit der Inhalte.

3 Erste Arbeiten: Sensordaten und Lernszenarios

Am Anfang des Vorhabens stand die Einrichtung eines grundsätzlich einsetzbaren Systems. Die Grundlage bildet der CBA-ItemBuilder, der für die Zwecke von SensoMot erweitert wird, um insbesondere die Integration und Auswertung der Sensordaten in nahezu Echtzeit und die Adaption der Lerninhalte zur Laufzeit zu ermöglichen. Zur Steuerung der Inhaltsadaption durch einen Motivationsindikator bedarf es zum einen der Integration geeigneter Sensoren und zum anderen der Adaption geeigneter Inhalte.

Für SensoMot wurden Sensoren aus Usability- und aus technischer Sicht auf ihre Eignung für das Projekt überprüft. Fünf Wearables, eine Smartwatch und vier Fitness-Tracker, wurden in Form einer Vorstudie qualitativ mit Studierenden evaluiert. Die Vorstudie setzte sich aus mehreren Usability Testdurchläufen inklusive Self-Reports, einer Anforderungsanalyse und Fokusgruppen mit Studierenden zusammen. Die Usability Tests basierten auf einer funktionalen Analyse zur Identifikation von Benchmark Tasks [HP12]. So konnten repräsentative Benutzeraufgaben kreiert werden, bei denen die Performance, die Bedienbarkeit und die User Experience der Geräte im Vordergrund standen. Darüber hinaus wurden die Effizienz und Effektivität der Geräte als zentrale Usability Metriken [TA08] evaluiert. Generell wurde die Usability der Geräte gut bis sehr gut bewertet. Die User Experience der Smartwatch übertrifft jene der Fitness-Tracker, insbesondere wegen größerer Funktionalität und dem Vorhandensein eines Displays zur Visualisierung des Mensch-Maschine Dialogs. Für eine Mehrheit der Befragten würden Wristband-Wearables bisher nur eine Nice to Have Ergänzung zu ihrem Smartphone darstellen [Mi15]. Trends wie Quantified Self als Katalysator eines neuen Bewusstseins der auf Datenanalyse basierenden Selbstreflexion [Ch14] in Verbindung mit dem Aufzeigen eines echten Mehrwerts der Anwendung für spezifische Anwendungsfälle könnten die Akzeptanz der Benutzer jedoch deutlich steigern.

Die Sensoren für eine Integration in SensoMot direkt ansprechbar sein, damit die Einzelmessungen in Echtzeit auswertbar und nicht vorverarbeitet sind. Außerdem wäre der Datenabfluss über eine Cloud aus datenschutzrechtlicher Sicht problematisch. Leider konnten nur äußerst wenige Wearables gefunden werden, welche die Sensordaten nicht über die Server des Herstellers leiteten – und die meisten davon waren sehr schnell nicht mehr erhältlich. Daher ist zusätzlich die Entwicklungsplattform MySignals auf ihre Eignung für das Projekt getestet worden: Der Zugriff auf die Sensoren kann unmittelbar durch Programmierung der Firmware erfolgen, jedoch ermöglichen Bauform und die prototypischen Steckverbindungen zur Anbindung der Sensoren keinen unproblematischen Feldeinsatz.

Zur Integration der Sensoren in den CBA-ItemBuilder wurde die Adaption von Lerninhalten in einen generischen Webservice ausgelagert, um mit verschiedenen Adaptionen algorithmen experimentiert zu können. Eine für Android entwickelte SensoMot App integriert die verschiedenen per Bluetooth gekoppelten „Wearables“, indem sie die Datenströme der unterschiedlichen Sensoren sammelt und zur Auswertung mittels Motivationsindikatoren weiterleitet. Diese werden mit Methoden aus dem Bereich des maschinellen Lernens [Bi06] bestimmt. Insbesondere wird der Ansatz des überwachten Lernens verfolgt, bei dem aus Beispielen von Sensordaten-Paaren und erhobenen Zielvariablen eine Funktion von den Sensordaten auf diese Zielvariablen gelernt wird.

Um SensoMot in verschiedenen Szenarios testen zu können, werden Inhalte für zwei unterschiedliche Themen und Ausbildungsarten entwickelt: Zum einen ein Inhaltssegment, das dem Lernziel „Grundlegende elektrotechnische Zusammenhänge verstehen“ [St12] des Fernlehrgangs mit Präsenzphasen zum staatl. geprüften Techniker (m/w) für Maschinenbautechnik folgt. Das Lehrgangziel wird durch fachsystematische und hand-

lungsorientierte Bestandteile wie fächerübergreifende Handlungsaufgaben [Ot11] in einem berufsbegleitenden Fernlehrgang umgesetzt. Da der lineare Lehrbrief in Papierform das zentrale Medium ist, wird dieser in eine adaptive Form mit mehrschichtigem fernstudienaffinen Inhalt überführt, um verschiedene Lernpfade anbieten zu können [Gr13]. Zum anderen erfolgt die Entwicklung einer adaptiven eLearning Plattform zur Unterstützung der ingenieurwissenschaftlichen Lehre. Hierfür wurde die Bachelorvorlesung Mikrotechnik gewählt, um die Problematik hoher Studienabbruchquoten in ingenieurwissenschaftlichen Studiengängen zu adressieren [He15]. Aktuell werden User Story Maps [PE15], Use Cases und Benutzerszenarios [HP12] erstellt, die in erste Prototypen, User Models [BM07] und Adaptionstechniken münden sollen.

4 Erste Schlussfolgerungen

Die Usability und User Experience aktueller auf dem Markt verfügbarer Wearables haben sich in Nutzertests als gut bis sehr gut erwiesen. Die Akzeptanz und Anwendungspotenziale der Technologie fokussieren nach Meinung der Nutzer besonders auf den Fitness- und Lifestyle-Aspekt der Geräte. Die Überprüfung der Eignung und Akzeptanz im Lernkontext steht noch aus und wird von der sinnvollen Einbettung der Technologie in den Lernprozess abhängen. Derzeit besteht jedoch das Problem, dass kein uns bekannter marktüblicher Sensor den Anforderungen des Projekts genügt. Obwohl der Markt an Wearables sehr stark wächst, gehen diejenigen Produkte mit offenen Schnittstellen eher zurück.

Bei der Erstellung des Adaptionsschemas für die berufliche Weiterbildung in Fernlehre besteht die Herausforderung darin, dass in der Entwicklungsphase bereits grundlegende didaktische und methodische Entscheidungen getroffen werden müssen. Deswegen müssen möglichst früh Dozenten- und Lerner-Einstellungen evaluiert werden und die Ergebnisse in die Entwicklung des Adaptionsschemas rückfließen.

5 Danksagung

Das Projekt SensoMot wird vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Förderschwerpunktes „Erfahrbares Lernen“ gefördert.

Literaturverzeichnis

- [Bi06] Bishop, C. M.: Pattern recognition and Machine Learning. Springer Science + Business Media LLC, 2006.
- [BM07] Brusilovsky, P.; Millán, E.: User Models for Adaptive Hypermedia and Adaptive Educational Systems. In (Brusilovsky, P.; Kobsa, A.; Nejd, W., Hrsg.): The Adaptive Web – Methods and Strategies of Web Personalization, Jgg. 4321 in Lecture Notes in

- Computer Science, S. 3–53. Springer, Berlin, Heidelberg, 2007.
- [Ch14] Choe, E. K.; Lee, N. B.; Lee, B.; Pratt, W.; Kientz, J. A.: Understanding Quantified-Selfers’ Practices in Collecting and Exploring Personal Data. In: CHI ’14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, S. 1143–1152, April 2014.
- [Gr13] Grassl, R.: Lehren an der Tastatur – Autorinnen und Autoren von Studienheften als Lehrende in der Distance Education. In (Cendon, E.; Grassl, R.; Pellert, A., Hrsg.): Vom Lehren zum Lebenslangen Lernen: Formate akademischer Weiterbildung, S. 117–133, Waxmann, Münster, New York, NY, München, Berlin, 2013.
- [He15] Heublein, U.; Ebert, J.; Hutzsch, C.; Isleib, S.; Richter, J.; Schreiber, J.: Studienbereichsspezifische Qualitätssicherung im Bachelorstudium – Befragung der Fakultäts- und Fachbereichsleitungen zum Thema Studierenerfolg und Studienabbruch, Jgg. 3 in Forum Hochschule. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, Hannover, August 2015.
- [HP12] Hartson, R.; Pyla, P. S.: The UX Book – Process and Guidelines for Ensuring a Quality User Experience. Elsevier, Morgan Kaufmann, Amsterdam, März 2012.
- [HSC07] Hektner, J. M.; Schmidt, J. A.; Csikszentmihalyi, M.: Experience Sampling Method – Measuring the Quality of Everyday Life. SAGE Publications, 2007.
- [Ma12] Martens, T.: Was ist aus dem Integrierten Handlungsmodell geworden? In (Kempf, W.; Langeheine, R., Hrsg.): Item-Response-Modelle in der sozialwissenschaftlichen Forschung, S. 210–229, Regener, Berlin, 2012.
- [Mi15] Min, C.; Kang, S.; Yoo, C.; Cha, J.; Choi, S.; Oh, Y.; Song, J.: Exploring Current Practices for Battery Use and Management of Smartwatches. In: Proceedings of the 2015 ACM International Symposium on Wearable Computers. ISWC ’15, Grand Front Osaka, ACM, New York, NY, USA, S. 11–18, September 2015.
- [Ot11] Ott-Kroner, A.: Umsetzung der Projektmethode im Fernlehrgang. Erziehungswissenschaft und Beruf, 59/01, S. 103–106, 2011.
- [PE15] Patton, J.; Economy, P.: User Story Mapping – Die Technik für besseres Nutzerverständnis in der agilen Produktentwicklung. O’Reilly, Sebastopol, April 2015.
- [Re05] Reinmann, G.: Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung. Unterrichtswissenschaft: Zeitschrift für Lernforschung, 33/01, S. 52–69, 2005.
- [Rö12] Rölke, H.: The ItemBuilder: A Graphical Authoring System for Complex Item Development. In (Bastiaens, T.; Marks, G., Hrsg.): Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012. Association for the Advancement of Computing in Education (AACE), Chesapeake, VA, S. 344–353, 2012.
- [St12] Staatsinstitut für Schulqualität und Bildungsforschung, München. Lehrpläne für die Fachschule für Maschinenbautechnik 7 2012, http://www.isb.bayern.de/download/15832/lp_fs_maschinenbautechnik_entwurf.pdf, Stand: 01.03.2017.
- [TA08] Tullis, T.; Albert, B.: Measuring the User Experience – Collecting, Analyzing, and Presenting Usability Metrics. Elsevier, Morgan Kaufmann, Amsterdam, 2008.

Generalisierbarkeit von Gamification-Ansätzen in E-Learning – eine explorative Studie

Isabel Slawik¹

Abstract: Gamification gewinnt als Ansatz zur Steigerung der Motivation und Partizipation in der Lehre zunehmend an Aufmerksamkeit, jedoch lassen bisherige heterogene Studienergebnisse Zweifel an der universellen Einsetzbarkeit des Ansatzes aufkommen. In einer explorativen Studie wurden die Einflussfaktoren eines freiwilligen Wettbewerbs auf Leistung und Motivation von Informatik-Studierenden untersucht, sowie der Einfluss von Spielertypen im Lehrkontext analysiert. Die Verknüpfung von Verhaltensdaten mit Aussagen aus strukturierten Interviews zeigte die bisher kaum berücksichtigte Notwendigkeit sowohl einer Individualisierung von Gamification-Maßnahmen als auch einer Adaption auf situative Veränderungen auf. Darüber hinaus wurden Möglichkeiten zur Nutzung von Gamification als Rückmeldung an den Dozenten identifiziert.

Keywords: Gamification, Lehre, Spielertypen, Individualisierung, Adaption.

1 Einleitung

Im Rahmen zunehmender Studierendenzahlen an deutschen Universitäten, insbesondere im MINT-Bereich, ergeben sich neue Herausforderungen für die Gestaltung von Vorlesungen mit mehreren hundert Teilnehmern. Online-Lernplattformen bieten die Möglichkeit für einen breiten Benutzerkreis Lehrmaterialien bereitzustellen sowie computergestütztes Arbeiten und Kollaboration zu fördern [KS15]. Nach einer Analyse von Piotrowski [Pi09] kann ein Großteil der bisher eingesetzten Lernmanagementsysteme (LMS) als ein Dokumentenservice (CMS-System) abgebildet werden, der um E-Learning-spezifische Assessment-Komponenten, bestehend aus Auswertungs- und Feedback-Komponenten, erweitert wurde. Diese Modularität erlaubt es häufig auf einfachem Wege bestehende Funktionalität durch neue Komponenten zu erweitern. So bietet beispielsweise das weitverbreitete LMS Moodle² ein umfangreiches Plug-in-Verzeichnis an. Unter anderem findet sich dort eine Vielzahl von Gamification-Plug-ins, die Kurse um spielerische Elemente wie Erfahrungspunkte (XP), Badges oder Ranglisten erweitern. Gamification beschreibt dabei die Anwendung von Spielelementen auf spielfremde Kontexte [De11]. Im Lernkontext erhofft man sich davon eine Steigerung der Motivation der Lernenden sowie positive Einflüsse auf das Verhalten [Gr15].

Die motivationale Wirkung von Spielen ist hinlänglich bekannt und im pädagogischen Umfeld bereits vielfach eingesetzt [Mc12]. Bisherige Ansätzen im Rahmen von „game-based learning“ oder „serious games“ konzentrierten sich darauf, Lerninhalte innerhalb

¹ LMU München, Institut für Informatik, Oettingenstraße 67, 80538 München, i.slawik@gmail.com

² <https://moodle.org>

von Spielen zu vermitteln. Im Gegensatz dazu ist Gamification eine Anwendung von Spielmechaniken unabhängig von dem zugrundeliegenden Lehrmaterial [LA14]. Landers [LA14] betont, dass mithilfe von Gamification lediglich eine Auswirkung auf das Verhalten und die Einstellung von Lernenden erzielt werden kann, nicht jedoch auf Lernfortschritte, die maßgeblich vom Lehrinhalt abhängen. Dies ist insofern zielführend, als dass aktive Beteiligung, kontinuierliche Mitarbeit und schnelle Rückmeldung als wirksame Faktoren in der Lehre empirisch belegt sind [vgl. Sc14b, S.201 ff]. Da Gamification immer nur eine Erweiterung oder Unterstützung des bestehenden Lehrinhalts ist, verspricht der Ansatz eine hohe Generalisierbarkeit auf verschiedenste Inhalte und Kontexte sowie eine einfache Integration in bereits bestehende Lehrkonzepte.

Damit stellt sich die Frage, ob einzelne Elemente aus einer anderen Welt – hier der Spielwelt – auf den Lernkontext übertragbar sind und inwieweit sich dadurch zusätzliche Potenziale des E-Learning erschließen lassen. Bislang gibt es dazu noch wenig aussagekräftige Befunde, obgleich sich dieses Vorgehen in letzter Zeit einer steigenden Popularität erfreut. So reichen Gamification-Ansätze von der Hinzunahme einzelner Spielelemente wie Badges [BS14] bis hin zu vollständig selbstorganisierten Kursen [BZ13]. Die Annahme der Übertragbarkeit der motivationalen Auswirkungen dieser Elemente erscheint zunächst plausibel, jedoch existieren vielfach keine ausreichenden empirischen Belege für die konkrete Wirkungsweise einzelner Spielelemente in Lernkontexten. Zwar gibt es erste Bestrebungen geeignete Messinstrumente für die Wirksamkeit von Gamification-Ansätzen zu entwickeln [KHG15] und differenzierte Modelle für die Kombination von Faktoren zu erstellen [Sc14a], doch sind dabei die Fragen, welche Faktoren überhaupt eine sinnfällige Kombination ergeben und ob diese Wirkungen unabhängig vom Einsatzkontext sind, noch nicht ausreichend erschlossen. Hinzu kommt, dass der Stand der Forschung in Bezug auf die Evaluation von Gamification-Ansätzen ein heterogenes Bild liefert [HKS14, SF15]. Zwar fanden quantitative Studien häufig eine positive Tendenz von Gamification-Maßnahmen, teilweise blieben erwartete Effekte jedoch aus [HKS14] oder waren gemischt positiv und negativ [SF15]. Ebenso wurde manches, was von einigen Benutzern in qualitativen Studien als besonders positiv hervorgehoben wurde, von anderen wiederum als besonders störend empfunden [HKS14]. So ergibt sich trotz einer Vielzahl von Fallstudien bis heute kein klares Bild davon, welche Spielelemente welche Wirkung zur Folge haben oder haben könnten. Schließlich muss noch festgehalten werden, dass die Frage, ob sich bekannte individuelle Präferenzen in Spielen und darauf aufbauende Typisierungen von Spielern auch auf den Einsatz von Spielelementen in spielfremden Kontexten übertragen lassen, ebenfalls noch weitgehend offen ist. So finden Konert et al. [Ko14] statistisch einige Überdeckungen zwischen Spielertypen und allgemeinen Persönlichkeitsfaktoren sowie der Typisierung von Lernenden, jedoch reichen die Befunde für die effektive Gestaltung von Gamification-Mechanismen in einer Lernplattform nicht aus.

Diese Befunde lassen die Annahme, die Wirkungsweise von Spielelementen ließe sich eins zu eins auf spielfremde Kontexte übertragen, problematisch erscheinen. Ebenso liegt die Vermutung nahe, dass Störfaktoren wie der Kontext, die genaue Auswahl von Elementen oder individuelle Vorlieben die Wirkungsweise von Gamification-

Maßnahmen beeinflussen. Die vorliegende Arbeit versucht, sich mit einer explorativen Studie in sehr kontrollierter Umgebung mithilfe intensiver Befragung diesen Fragen zu nähern und die Auswirkung eines einzelnen Spielelements auf die Verhaltensweise und Motivation von Lernenden zu analysieren. Mit der Absicht, die dahinterstehenden Fragen genauer ins Blickfeld zu nehmen, wurde der Fokus explizit auf den explorativen Charakter der Studie gelegt, um eine weitere Fallstudie mit wechselnder Konstellation von Spielelementen, Kontext und Einflüssen zu vermeiden. Ziel waren die Erörterung der Einflussfaktoren auf die Motivation von Lernenden sowie die Analyse des Einflusses von Spielertypen auf Lernende in einem Gamification-Kontext.

2 Vorgehen

Die Studie wurde im Rahmen eines Softwareentwicklungspraktikums an der LMU München durchgeführt. Die Veranstaltung ist Teil des Pflichtstudiums im dritten Semester der Bachelorstudiengänge Informatik und Medieninformatik, wobei verschiedene Praktika zur Auswahl standen. Im Praktikum entwickelten die Teilnehmer in Gruppen von je fünf Personen ein Multiplayer-Jump'n'Run-Spiel in JavaScript. Insgesamt standen 15 Plätze zur Verfügung. Als Gamification wurde ein Wettbewerb durchgeführt. Dieser basierte auf zusätzlichen freiwilligen Aktivitäten im Rahmen des Praktikums, mit dem Ziel den Lernprozess für die Projektarbeit zu unterstützen. Um den Einfluss von Spielertypen auf die Wirkung von Gamification in einem Lehrkontext zu untersuchen, wurde die Taxonomie von Spielertypen nach Richard Bartle [Ba96] benutzt. Diese Typisierung in Killer, Achiever, Socializer und Explorer ist nicht die einzig mögliche in diesem Feld, jedoch weiterhin eine der populärsten und ermöglicht somit die Anschlussfähigkeit zu anderen Untersuchungen wie z. B. von Konert et al. [Ko14].

Die geringe Teilnehmerzahl erlaubt zwar keine empirisch stichhaltigen Ergebnisse, war jedoch für den explorativen Charakter der Studie unabdinglich, um eine intensive Befragung der Studierenden zu ermöglichen. Verhaltens- und Leistungsdaten wurden anhand von Protokollen ausgewertet und durch strukturierte Interviews in Kontext gesetzt.

2.1 Ablauf

Die Teilnehmer des Praktikums bekamen vor Beginn der Lehrveranstaltung einen Online-Fragebogen zugesandt. Dieser bestand aus drei Teilen: einem Teil zu bisherigen Programmierkenntnissen, einem Teil zu Spielverhalten und -vorlieben sowie abschließend dem Bartle-Test for Gamer Psychology [AD00]. Die Reihenfolge der Antworten sowie der jeweiligen Antwortmöglichkeiten für den Bartle-Test wurden für jeden Teilnehmer randomisiert. Das Ausfüllen des Fragebogens war die erste Aufgabe des Wettbewerbs.

Der Wettbewerb fand in insgesamt 9 Wocheneinheiten statt und dauerte vom Vorlesungsbeginn im Oktober bis Ende Dezember. Um die Teamarbeit zu fördern und soziale Einflussfaktoren zu ermitteln, wurden sowohl ein Gruppen- als auch ein Einzelwettbe-

werb durchgeführt. Die Wettbewerbsgruppen entsprachen den Projektarbeitsgruppen. Die Teilnahme am Wettbewerb war freiwillig, ein Ausscheiden war jederzeit ohne persönliche negative Konsequenzen möglich. Im Gegensatz zu einigen Ansätzen aus der Literatur [Be15, SSS14] wurde hier bewusst auf eine Kopplung von Gamification-Element an Noten der Lehrveranstaltung verzichtet, um den motivationalen Einfluss des Wettbewerbs isoliert betrachten zu können. Als Teilnahmeanreiz wurde ein Gruppenpreis (Pizza & Getränke für die Gruppe) sowie ein Einzelpreis (20€ Steam³-Gutschein) ausgeschrieben. Da die Teilnahme am Wettbewerb eine freiwillige Mehrarbeit für die Studierenden bedeutete, wurde darauf geachtet, die Aufgaben an den Lernzielen des Praktikums (Erlernung von Programmierparadigmen anhand von JavaScript, Entwicklung eines Spiels im Rahmen einer Gruppenarbeit) auszurichten.

Studierende konnten in wöchentlichen Einzel- und Gruppenaufgaben Punkte für den Wettbewerb sammeln. Einzelaufgaben waren Programmieraufgaben in JavaScript, die nach Korrektheit, Effizienz und Stil bewertet wurden. Gruppenaufgaben orientierten sich an der Projektarbeit und wurden in wöchentlichen Tutorentreffen besprochen. Zusätzlich wurden unregelmäßig Bonusaufgaben (Gruppen- oder Einzelaufgaben) angeboten, die weniger eng mit den Lernzielen des Praktikums verknüpft waren.

Punkte aus den Einzelaufgaben wurden zu einer persönlichen Punktzahl addiert. Für den Gruppenwettbewerb wurden die Punktzahlen aller Teammitglieder mit den Punkten der Gruppenaufgaben summiert. Im wöchentlichen Plenum wurde eine Rangliste der Gruppen, sowie die beste Einzelpunktzahl (anonym) präsentiert. Zusätzlich erhielt jeder Studierende eine wöchentliche Übersicht über die eigenen Punkte je Aufgabe, die Gesamtpunktzahl jedes Gruppenmitglieds und die aktuelle Rangliste der Gruppen per E-Mail.

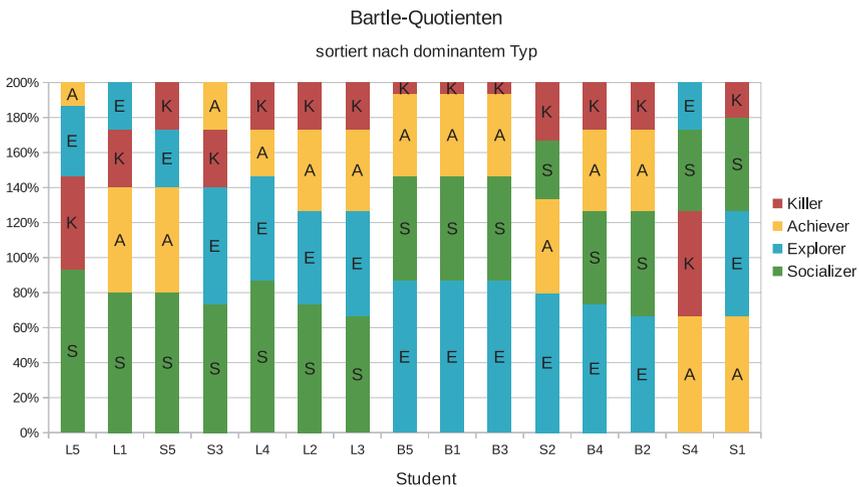


Abb. 1: Bartle-Quotienten für Teilnehmer des Praktikums (anonymisiert)

³ <http://store.steampowered.com>

3 Studienverlauf

Am Praktikum nahmen 15 Studierende teil, alle Plätze waren belegt. Jeder Teilnehmer füllte den Online-Fragebogen vor Veranstaltungsbeginn aus. Die Studierenden waren im 3. bis 5. Semester und zwischen 19 und 23 Jahren alt. Das Geschlechterverhältnis war ausgeglichen mit 7 weiblichen und 8 männlichen Studenten. Alle gaben an, Grundkenntnisse bis gute Programmierkenntnisse und keine bis Grundkenntnisse in JavaScript zu besitzen. Für die meisten war das Praktikum das erste große Programmierprojekt außerhalb der Anfängervorlesungen.

3.1 Auswertung der Protokolle

Die Studierenden wurden anhand der Ergebnisse des Bartle-Test of Gamer Psychology in drei Gruppen zu je fünf Teilnehmern eingeteilt. Abbildung 1 zeigt die errechneten Bartle-Quotienten, sortiert nach den dominanten Spielertypen. Rund die Hälfte der Studierenden identifizierte sich als Socializer (46,67%), die Mehrheit der verbleibenden als Explorer (40%). Auffällig ist, dass nur zwei Studierende vom Typ Achiever (13,33%) dabei waren und keiner in die Kategorie Killer (0%) als dominantem Typ fiel. Eine ähnliche Verteilung ergab sich in einer Vergleichsumfrage unter Teilnehmern der parallel angebotenen Praktika. Von 29 Teilnehmern waren Explorer (58,62%) der am häufigsten dominante Typ, gefolgt von Socializern (20,69%). Achiever (6,9%) und Killer (3,45%) waren nur gering vertreten. Drei Teilnehmer konnten keinem dominanten Typ zugeordnet werden, da sie zu gleichen Anteilen Achiever und Socializer bzw. Explorer waren.

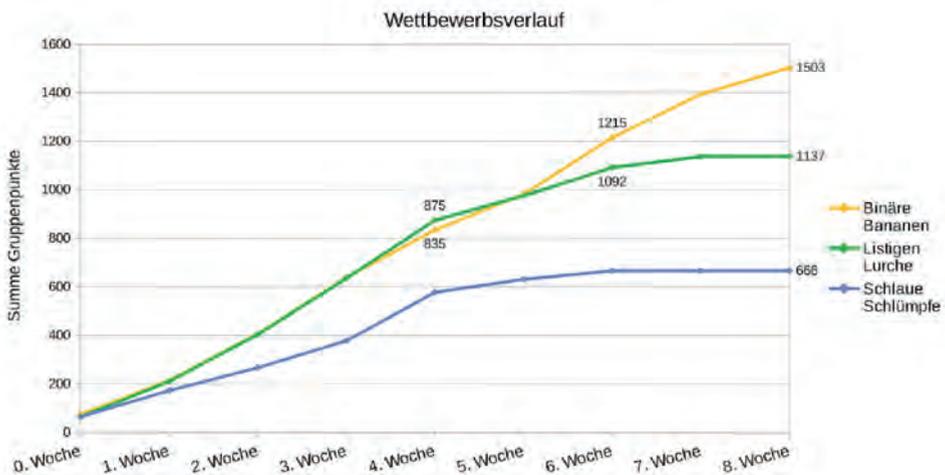


Abb. 2: Gesamtpunkte der Gruppen über den Verlauf des Wettbewerbs

Die Teilnehmer des Praktikums wurden in eine Socializer-Gruppe („Listigen Lurche“, Studenten L1-L5 in Abb. 1), eine Explorer-Gruppe („Binären Bananen“, Studenten B1-B5) und eine Gruppe mit den höchsten Achiever- und Killer-Anteilen („Schlaue Schlümpfe“, Studenten S1-S5) eingeteilt. Dabei wurde darauf geachtet, die Gruppen in Geschlecht, Studiengang und Programmierkenntnissen möglichst ausgeglichen zu gestalten. Die Gruppennamen wurden von den Gruppen selbst frei gewählt.

Abbildung 2 zeigt den Verlauf des Gruppenwettbewerbs. Zwei Mitglieder der Gruppe „Schlaue Schlümpfe“ beteiligten sich nicht wesentlich am Wettbewerb, dementsprechend blieb die Gruppe in ihrer Punktzahl weit hinter den anderen Gruppen zurück. Die Schlaue Schlümpfe stiegen insgesamt als erstes aus dem Wettbewerb aus, nach der 6. Woche nahm kein Mitglied mehr am Wettbewerb teil. Die Gruppe „Binäre Bananen“ war die ersten 4 Wochen stets einige Punkte besser als die Gruppe „Listigen Lurche“. In der 4. Woche übernahmen die Listigen Lurche kurzzeitig die Führung, fielen jedoch bereits in der darauffolgenden Woche wieder hinter die Binären Bananen zurück, welche letztendlich den Gruppenwettbewerb für sich entschieden.

Einzelsieger wurde ein Mitglied der Binären Bananen, welches als einziger Teilnehmer des Praktikums alle 17 individuellen Aufgaben bearbeitete. Im Mittel wurden 13 individuelle Aufgaben abgegeben. Die höchste persönliche Gesamtpunktzahl lag bei 320 Punkten, die niedrigste bei 29, wobei der Durchschnitt der Einzelpunktzahlen bei 172,07 Punkten lag (Median 176). Weder in der Anzahl der bearbeiteten Einzelaufgaben noch in der Punktzahl ergaben sich nennenswerte Unterschiede zwischen den Geschlechtern.

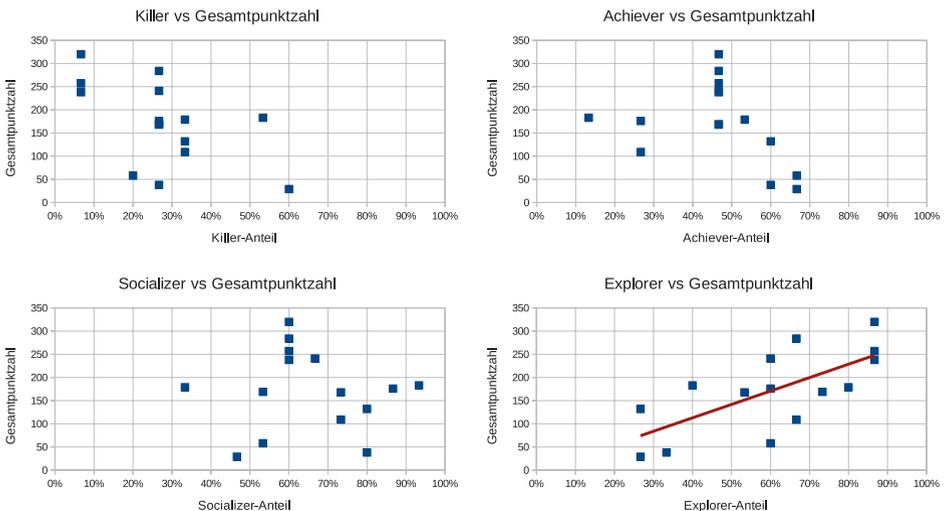


Abb. 3: Verhältnis von Bartle-Spielertypen zu persönlichen Gesamtpunktzahlen

In Abbildung 3 sind die Prozentsätze der Bartle-Spielertypen aus den Ergebnissen des Bartle-Test of Gamer Psychology im Verhältnis zur Einzelpunktzahl am Ende des Wettbewerbs aufgezeichnet. Jedes Diagramm enthält alle 15 Studierenden. Die rote Linie zeigt eine statistisch signifikante Korrelation ($r = 0,69$, $p < 1\%$) für den Explorer-Spielertyp mit der Gesamtpunktzahl an. Dies bedeutet, dass ein höherer Bartle-Quotient im Explorer-Typ mit einer höheren Gesamtpunktzahl im Wettbewerb einherging.

3.2 Auswertung der Interviews

Die Klassifikation in Spielertypen nach dem Bartle-Test stimmte im Großteil mit der Selbsteinschätzung der Studierenden überein. Danach gefragt, in welchem Bartle-Typ sie sich am ehesten wiedererkennen, gab die Mehrheit (60%) eine Antwort, die mit ihrem dominanten oder ihren zwei dominantesten Typen nach dem Bartle-Test übereinstimmte. Drei Studierende (20%) gaben einen abweichenden Typ an und weitere drei Studierende (20%) nannten zwei Typen, von denen einer mit ihrem dominanten Typ nach dem Bartle-Test übereinstimmte, der zweite jedoch nicht.

Obwohl zwei Mitglieder der Schlaunen Schlümpfe nicht am Wettbewerb teilnahmen, gaben die übrigen Mitglieder in den strukturierten Interviews an, dass sie dies nicht gestört hätte. Trotzdem bleibt festzustellen, dass die Gruppe am frühesten aus dem Wettbewerb ausschied. Im Gegensatz dazu zeigten die Gruppen „Listigen Lurche“ und „Binäre Bananen“ ein starkes Sozialgefüge innerhalb der Gruppe. Obwohl niemand negative Konsequenzen zu fürchten hatte, war die Teilnahme der übrigen Gruppenmitglieder ein starker Anreiz für Einzelne ebenfalls mitzumachen, wie durch die folgende Aussage beispielhaft belegt wird: *„Wenn die ganze Gruppe das macht, macht man halt einfach mit um zu sagen ‚ja, ich tue auch was dafür‘.“* Das Mitglied der Binären Bananen, welches am wenigsten Einzelaufgaben bearbeitete, hat immer noch genauso viele Aufgaben absolviert wie die zwei Mitglieder der Schlaunen Schlümpfe, die am meisten Aufgaben aus der Gruppe abgegeben haben. Insofern lässt sich ein deutlicher Unterschied in der Motivation der Gruppen feststellen.

Naiv könnte man annehmen, dass Achiever und Killer stärker durch eine Wettbewerbsdynamik motiviert werden als Socializer oder Explorer. Dies konnte in der Studie nicht bestätigt werden. Tatsächlich war die Gruppe „Schlaue Schlümpfe“ die am wenigsten am Wettbewerb beteiligte Gruppe, obwohl ihre Mitglieder die stärksten Achiever- und Killer-Anteile besaßen. Der Student, der den stärksten Killer-Quotient aufwies und sich selbst als Killer einschätzte, hat am Wettbewerb von allen Studierenden am wenigsten Aufgaben abgegeben. Obwohl er Zweikämpfe in Spielsituationen motivierend findet, war der Wettbewerb kein Anreiz für ihn teilzunehmen. Jemanden direkt im Wettbewerb zu besiegen *„...wäre jetzt nicht so Anreiz gewesen. Wenn dann muss ich schon direkt beim Spiel etwas besiegen. Das ist ja wieder was Anderes.“*

Ebenso kann man nicht, wie häufig implizit angenommen, davon ausgehen, dass ein

Spielelement eine konstante Wirkung auf die Motivation von Studierenden entfaltet. Dies wird besonders deutlich am Beispiel der Gruppe „Listigen Lurche“. Die Gruppenmitglieder waren anfangs stark motiviert am Wettbewerb teilzunehmen, insbesondere der enge Zweikampf mit den Binären Bananen war ein Ansporn. Nach ihrem Triumph in der 4. Woche schied jedoch ein Mitglied der Listigen Lurche aus dem Wettbewerb aus. Dies sorgte für Unmut unter den Teammitgliedern. Bestrebungen innerhalb der Gruppe das Mitglied zur weiteren Teilnahme am Wettbewerb zu motivieren scheiterten jedoch. In der Folge fielen die Listigen Lurche in der 6. Woche so weit hinter die Binären Bananen zurück, dass ein Gruppensieg unerreichbar erschien. Daraufhin entschied die Gruppe, gemeinsam aus dem Wettbewerb auszusteigen. Lediglich ein Mitglied gab in der 7. Woche noch eine Aufgabe ab. Ebenso wurde auf individueller Ebene berichtet, dass Motivation in Demotivation umschlug, wenn der Einzelsieg unerreichbar erschien. Dass ein Wettbewerb auch demotivierend wirken kann ist zwar hinlänglich bekannt, jedoch fanden dynamische Motivationseffekte in der Untersuchung von Gamification-Maßnahmen bisher keine Berücksichtigung in der Literatur.

Die Gründe am Wettbewerb teilzunehmen waren vielfältig. Rund die Hälfte der Studierenden gab an, die Programmieraufgaben bearbeitet zu haben, da sie beim Lernen geholfen hätten. *„Weil ich lernen wollte“* oder *„Hauptsächlich habe ich es gemacht, dass ich mich in JavaScript einarbeiten kann“* waren Aussagen aus den strukturierten Interviews, die dies belegen. Zwei Studierende gaben ein Vertrauen in die Lehrperson als Grund an: *„Meistens ist es ja so, dass es auch mit Hintergrund gemacht wird sowas, und deswegen versucht man ja auch da irgendwie mal zumindest am Anfang noch dabei zu bleiben.“* Als zweithäufigster Grund wurde die soziale Verpflichtung gegenüber dem Team genannt, so zum Beispiel ein Mitglied der Listigen Lurche: *„Ich hab’ eh hauptsächlich mitgemacht, weil es halt blöd fürs Team ist, wenn man es schleifen lässt.“* Weiterhin wurden Spaß an den Aufgaben, ein klares Ziel auf das man hinarbeiten konnte, die Herausforderung an einen selber und der Wettbewerb als Ansporn genannt. Ein Student fasste es so zusammen: *„Anfangs sicherlich auch noch der Spielaspekt, so der Wettbewerbsaspekt, das hat aber relativ schnell abgenommen, weil ich dann irgendwo auch an den Aufgaben selbst Spaß dran hatte.“* Eine Studentin sagte, *„am Anfang ich habe es für [die] Übung gemacht, [...] dann irgendwo, wo ich nicht mehr so gut drangekommen bin, habe ich es irgendwie für [das] Team gemacht.“* Dies zeigt insbesondere, dass auch auf persönlicher Ebene Motivationsfaktoren situativ und dynamisch sein können.

Als Gründe für den Ausstieg aus dem Wettbewerb wurden überwiegend Zeitgründe genannt. Die Aufgaben wurden über den Verlauf des Wettbewerbs schwieriger, die Projektarbeit wurde anspruchsvoller, andere Vorlesungen nahmen mehr Zeit in Anspruch und am Ende des Jahres kamen die Feiertage hinzu. Viele Studierende gaben an, lieber die vorhandene Zeit in die Projektarbeit, welche in die Endnote einfluss, zu investieren, als in den freiwilligen Wettbewerb. Die Aussichtslosigkeit auf den Einzel- bzw. Gruppensieg wurde ebenfalls als demotivierender Faktor genannt. So gab ein Mitglied der Schlaun Schlümpfe, welches sich nicht nennenswert am Wettbewerb beteiligt hat, an: *„Ich war ein bisschen faul am Anfang und dann irgendwann war der Zeitpunkt, da waren die anderen so weit vorne, dass man auch nicht mehr nachgekommen ist.“* Das zwei-

te Mitglied der Schlaunen Schlümpfe, welches sich kaum beteiligte, gab einen fehlenden Ehrgeiz innerhalb der Gruppe als Grund an: „... und dann schienen irgendwie alle aus der Gruppe jetzt nicht so darauf aus zu sein, dass wir jetzt die Ersten werden.“

Obwohl sich deutliche Unterschiede in den Motivationslagen der Studierenden ausmachen lassen, fällt es schwer die Begründungen auf einzelne Spielertypen abzubilden. Vielmehr scheinen andere, noch unbekanntere Persönlichkeitsfaktoren eine Rolle zu spielen. Ebenso gab es Anhaltspunkte, dass bereits bestehende Verhaltensmuster im Lernkontext einen Einfluss auf das Verhalten der Studierenden im Wettbewerb hatten. Ein Student gab an, Übungsblätter erst gegen Ende der Vorlesung zur Klausurvorbereitung gebündelt zu bearbeiten und zeigte ein ähnliches Verhalten im Wettbewerb: „*Ich bin da allgemein eher so der ‚ich schau es mir hinterher an‘...*“ Ein Mitglied der Binären Bananen meinte, „*wie gesagt, ich krieg bei sowas extrem schnell ein schlechtes Gewissen, wenn ich sowas nicht mache.*“ Sie gab an, Übungsblätter ebenfalls stets sehr gewissenhaft während des Semesters zu bearbeiten und blieb beim Wettbewerb bis zum Ende engagiert. Hier kommen folglich Aspekte hinzu, die durch die vorherige Klassifikation nicht erfasst wurden und einer näheren Untersuchung bedürfen.

4 Diskussion

Darüber hinaus stellt sich die Frage, inwieweit eine statische Klassifikation der Studierenden ausreichend ist. Die Studie macht deutlich, dass neben individuellen Unterschieden auch dynamische Effekte eine wichtige Rolle spielen. So können sich die für die Motivation relevanten Faktoren im Laufe der Gamification-Maßnahme verschieben, wenn etwa die extrinsische Motivation durch ein Spielelement, hier der Wettbewerb, durch eine intrinsische Motivation, wie Spaß und Freude am Lernen, ersetzt wird.

Ebenso haben soziale Interaktionen einen starken Einfluss auf Einzelne, wie sich im Gruppenverhalten gezeigt hat. Während es den Binären Bananen gelungen ist, auch schwächere Mitglieder zur kontinuierlichen Teilnahme am Wettbewerb zu motivieren, hat der fehlende Ehrgeiz der Schlaunen Schlümpfe gemeinsam teilzunehmen dazu geführt, dass Mitglieder gar nicht am Wettbewerb teilnahmen oder frühzeitig aufgaben. Ferner zeigten sich hier, ebenso wie bei den persönlichen Motivationsfaktoren, dynamische Effekte. Das Verhalten einer einzelnen Person hatte einen destruktiven Effekt auf die übrigen Mitglieder. Obwohl die Listigen Lurche ursprünglich stark motiviert waren, bewirkte der Ausstieg eines Mitglieds die Demotivation der übrigen Gruppenmitglieder, so dass diese ebenfalls nicht weiter am Wettbewerb teilnahmen. Um solchen Effekten entgegenzuwirken ist eine Adaption von Gamification-Maßnahmen während ihres Verlaufs unabdinglich. Aus technischer Sicht wäre hier eine automatische Erfassung solcher Effekte durch geeignete Learning Analytics-Verfahren wünschenswert.

Dies führt auf einen weiteren Aspekt, der bisher in der Literatur kaum betrachtet wurde: die Möglichkeit zusätzlicher Rückmeldung an den Dozenten. Gamification könnte dazu genutzt werden, dem Lehrenden wichtige Informationen über die Motivationslage und

die Leistungen der Studierenden zu übermitteln. Dies würde den Dozenten befähigen, bei drohender Demotivation geeignete Maßnahmen zu ergreifen, z. B. die Gruppenzusammensetzung zu ändern, andere oder zusätzliche Anreize zu setzen, unterschiedliche Aufgaben anzubieten oder den Wechsel auf eine andere Spielmechanik einzuleiten.

Ebenfalls könnte eine Änderung in der Rückmeldung an die Studierenden selbst erfolgen. Obwohl bisher keine empirische Untersuchung dazu gefunden wurde, liegt die Vermutung nahe, dass das Motivationspotenzial eines Spielelements von der Art des Feedbacks abhängt. In der Studie waren die Studierenden an Feedback auf unterschiedlichen Ebenen interessiert. Während manche eine Highscore über alle Teilnehmer begrüßt hätten, standen andere dieser Idee neutral bis ablehnend gegenüber und bevorzugten individuelleres Feedback. Im Wettbewerb war durch die Punktemails und persönlichen Korrekturen der Programmieraufgaben Feedback auf drei Ebenen gegeben, die sich auf die drei Bezugsnormen von Rheinberg für die Bewertung schulischer Leistungen [Rh01] abbilden lassen: soziale (eigene Leistungen gemessen an den Leistungen anderer), individuelle (eigene Leistungen gemessen an den vorangegangenen eigenen Leistungen) und absolute (eigene Leistungen gemessen an absoluten Standards). Feedback auf Ebene der sozialen Bezugsnorm war im Praktikum durch die Angabe der Punktzahlen der Gruppenmitglieder in den wöchentlichen E-Mails gegeben. Auf individueller Ebene war die Auflistung der bisherigen Leistungen in den E-Mails gegeben und auf absoluter Ebene können die Kommentare und Bewertungen der Programmieraufgaben angesehen werden. In den Interviews gaben Studierende unterschiedliche Präferenzen für die jeweilige Feedback-Komponente im Praktikum und damit für die unterschiedlichen Bezugsnormen an, was auch hier für eine Individualisierung spricht. Nach Rheinberg [Rh01] ist bekannt, dass die Bezugsnormen auch unterschiedlich motivational wirken können. So ist für schwache Schüler eine individuelle Bezugsnorm, die die persönlichen Leistungssteigerungen hervorhebt, motivierender als ein sozialer Vergleich, in dem die Schüler stets einen der hinteren Plätze belegen. Insofern könnte auch im Rahmen von Gamification-Ansätzen eine Adaption der Bezugsnorm und der entsprechenden Rückmeldung an den Studierenden abhängig von den situativen Gegebenheiten angebracht sein.

5 Fazit

In der explorativen Studie wurde die implizite Annahme vieler Gamification-Ansätze, dass Gamification-Maßnahmen universell einsetzbar sind, genauer untersucht. Durch eine Verknüpfung von Verhaltensdaten mit Aussagen aus strukturierten Interviews wurde deutlich, dass ein einzelnes Spielelement wie ein Wettbewerb im Kontext einer Lehrveranstaltung sehr unterschiedlich auf Studierende wirken kann. Neben der Notwendigkeit zur Individualisierung von Gamification-Maßnahmen ergaben sich dynamische Effekte, welche für eine kontinuierliche Überprüfung der Motivationslagen und gegebenenfalls eine Adaption basierend auf Veränderungen in situativen Faktoren sprechen.

Die Abbildung von unterschiedlichen Motivationslagen auf Spielertypen und damit einer

direkten Übertragung der Motivationswirkung von Spielelementen innerhalb von Spielen auf Spielelemente in spielfremden Kontexten bleibt fragwürdig. Zwar fand sich eine positive Korrelation von Explorer-Spielertypen zur Gesamtpunktzahl im Wettbewerb, dies lässt sich jedoch durch die kleine Stichprobe und den Sieg der Explorer-Gruppe im Wettbewerb erklären. Die Gründe zur Teilnahme am Wettbewerb ließen keine eindeutige Abbildung auf Spielertypen zu. Vielmehr fanden sich Indikatoren, dass der Kontext, in dem Gamification durchgeführt wird, einen Einfluss auf das Verhalten hat. So traten im Wettbewerb bestehende Verhaltensmuster aus dem universitären Kontext zutage, die nicht durch Spielertypen begründbar sind. Hier würde es sich anbieten, Gamification auf Typisierungen zu beziehen, deren Wirksamkeit im universitären Lernalltag empirisch gesichert ist [vgl. 5-Klassen-Lösung in Sc14b]. Des Weiteren hat die Studie gezeigt, dass nicht nur die Auswahl von geeigneten Spielelementen, sondern auch eine differenzierte individuelle Rückmeldung an Studierende im Lernkontext eine wichtige Rolle einnimmt und gegebenenfalls über den Zeitraum einer Gamification-Maßnahme anzupassen ist.

Neben der Auswirkung auf Verhalten und Motivation von Studierenden bieten Gamification-Ansätze aber auch interessante Möglichkeiten für die Rückmeldung an Dozenten. So werden im Rahmen von Gamification typischerweise eine Vielzahl von Daten erfasst, die im Sinne von Learning Analytics ausgewertet werden könnten, um dem Lehrenden sowohl Hinweise über den Aktivitätsverlauf der Lernenden zu geben, als auch auf Basis von diesen Daten Indikatoren herauszufiltern, die eine Veränderung im didaktischen Vorgehen oder eine Änderung des Einsatzes von Spielelementen ermöglichen. Sofern dies gelingt, wäre es möglich, darauf basierend selbst-adaptive Gamification-Konzepte zu entwickeln, um negativen Auswirkungen von Gamification entgegenzusteuern.

Im Sinne einer hypothesengeleiteten Gestaltung bedarf es einer genauen Analyse, Theoriebildung und weiterer empirischer Untersuchungen im unmittelbaren Einsatzkontext, um eine Weiterentwicklung von Gamification-Ansätzen in Lernplattformen zielführend voranzutreiben. Eine statische Integration einzelner Spielelemente in Lernplattformen allein kann die in der Einleitung beschriebene Zielsetzung nur dann erreichen, wenn es gelingt die Vielfalt und Dynamik motivationaler Befindlichkeiten in einem entsprechenden didaktischen Szenario situativ zu berücksichtigen.

Literaturverzeichnis

- [AD00] Andreasen, E.; Downey, B.: Bartle Test of Gamer Psychology, 2000, <http://www.andreasen.org/bartle/>, Stand: 15. Juni 2017.
- [Ba96] Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit MUDs, *Journal of MUD research* 1/1, 1996.
- [Be15] Berkling, K.: Connecting Peer Reviews with Students' Motivation - Onboarding, Motivation and Blended Learning. In: 7th International Conference on Computer Supported Education, S. 24-33, 2015.
- [BS14] Buchem, I.; Sennewald, P.: Badges als Elemente von digitalen Lernumgebungen: Ein-

- satzszenarien am Beispiel von BeuthBadges. In: Die 12. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V., Freiburg, S. 169-180, 2014.
- [BZ13] Berkling, K.; Zundel, A.: Understanding the Challenges of Introducing Self-Driven Blended Learning in a Restrictive Ecosystem - Step 1 for Change Management: Understanding Student Motivation. In: Proc. of the 5th International Conference on Computer Supported Education, Aachen, S. 311-320, 2013.
- [De11] Deterding, S.; et al.: From game design elements to gamefulness: defining gamification. In: Proc. of the 15th International Academic MindTrek Conference, Tampere, Finland, S. 9-15, 2011.
- [Gr15] Grund, C.K.: How Games and Game Elements Facilitate Learning and Motivation: A Literature Review. In: 45. Jahrestagung der Gesellschaft für Informatik, Cottbus, S. 1279-1293, 2015.
- [HKS14] Hamari, J.; Koivisto, J.; Sarsa, H.: Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In: Proc. of the 47th Annual Hawaii International Conference on System Sciences, Waikoloa, USA, S. 3025-3034, 2014.
- [KHG15] Kettner, R.; Herrmann, K.; Gaulke, W.: Der IGQ – Ein Messinstrument für die Wirksamkeit von Gamification. In: Mensch und Computer Workshopband, Stuttgart, S. 359-365, 2015.
- [Ko14] Konert, J.; et al.: Modeling the Player: Predictability of the Models of Bartle and Kolb Based on NEO-FFI (Big5) and the Implications for Game Based Learning. *International Journal of Game-Based Learning*, 4/2, S. 36-50, 2014.
- [KS15] Keil, R.; Selke, H.: Virtuelle Wissensräume – Von der Präsentation von Inhalten zu virtuellen Lernstätten. In: 20 Jahre Lernen mit dem World Wide Web – Technik und Bildung im Dialog, Paderborn, Bd. 330, S. 39-53, 2015.
- [La14] Landers, R.: Developing a Theory of Gamified Learning Linking Serious Games and Gamification of Learning. *Simulation & Gaming* 45/6, S. 752-768, 2014.
- [Mc12] McClarty, K.L.; et al.: A Literature Review of Gaming in Education. Pearson, 2012.
- [Pi09] Piotrowski, M.: Document-Oriented E-Learning Components. Dissertation, Otto-von-Guericke-Universität Magdeburg, 2009.
- [Rh01] Rheinberg, F.: Leistungsbeurteilung im Schulalltag: Wozu vergleicht man was womit? In (Weinert, Franz E., Hrsg.): *Leistungsmessung in Schulen*, Beltz, 2001.
- [Sc14a] Schering, S.: Ein Design Space für interne Gamification-Anwendungen, In: Mensch und Computer 2014 Workshopband, München, S. 361-368, 2014.
- [Sc14b] Schulmeister, R.: Auf der Suche nach Determinanten des Studienerfolgs. In (Brockmann, J.; Pilniok, A., Hrsg.): *Studieneingangsphase in der Rechtswissenschaft*, 1. Auflage., Baden-Baden: Nomos, S. 72-205, 2014.
- [SF15] Seaborn, K. Fels, D.I.: Gamification in theory and action: A survey, *International Journal of Human-Computer Studies* 74, S. 14-31, 2015.
- [SSS14] Stöcklin, N.; Steinbach, N.; Spannagel, C.: QuesTanja: Konzeption einer Online-Plattform zur computerunterstützten Gamification von Unterrichtseinheiten. In: Die 12. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V., Freiburg, S. 151-156, 2014.

Empirische Untersuchung des EGameFlow eines Serious Games zur Verbesserung des Lernerfolgs

Linda Eckardt¹, Andreas Pilak¹, Manuel Löhr¹, Peer van Treel¹, Johannes Rau¹, Susanne Robra-Bissantz¹

Abstract: Die Integration von Spielelementen ist in der Lehre ein wichtiger Bestandteil geworden, um die Selbstinitiative des Anwenders zu steigern und damit den Lernerfolg zu verbessern. Das EGameFlow-Modell bietet die Möglichkeit das Spielerlebnis des Anwenders zu messen und Potentiale zur Verbesserung aufzuzeigen. Die vorliegende Studie untersucht daher den EGameFlow anhand eines Serious Games zum Lernen von Informationskompetenz. Im Vergleich zu bisherigen Studien wird dabei das EGameFlow-Modell mit allen Dimensionen in einem neuen Lernkontext betrachtet. Ergebnisse der Studie zeigen eine positive Bewertung der Dimensionen „Rückmeldung“, „Wissenszuwachs“, „klare Zieldefinierung“, „Herausforderung“, „Konzentration“ und „Autonomie“. Verbesserungspotentiale haben die Dimensionen „Immersion“ und „soziale Eingebundenheit“ gezeigt.

Keywords: EGameFlow, E-Learning, Serious Game, Lernerfolg, Informationskompetenz

1 Einleitung und Motivation

In den vergangenen Jahren ist die Bedeutung von E-Learning gestiegen. E-Learning verschafft dem Anwender die Möglichkeit sein Wissen autonom zu konstruieren. Dabei hängt der Lernerfolg stark von der Selbstinitiative des Anwenders ab [FSY09]. Damit diese gesteigert wird, ist die Integration von Spielelementen im E-Learning Umfeld, auch digitales Game-based Learning genannt, weit verbreitet [Pr01, RA03]. Es existieren verschiedene Ansätze zur Einbindung von Spielelementen in die Lehre. Unter Gamification wird die Integration einzelner Spielelemente in einen spielfremden Kontext verstanden, wohingegen ein Serious Game die Entwicklung eines vollwertigen Spiels mit festen Regeln und unter Berücksichtigung von Aspekten des Designs bezeichnet [De07]. Obwohl die Grenzen zwischen diesen beiden Ansätzen nicht eindeutig sind, haben sie doch etwas gemeinsam. Beide Ansätze integrieren Spielelemente, um die Lernenden zu motivieren sich aktiver und intensiver mit Themengebieten auseinanderzusetzen [De07].

Elemente wie Punkte sammeln, Missionen oder Formen eines Wettkampfes helfen dem Anwender beim Eintauchen in ein Spiel. Das lässt auf Zusammenhänge des Spielerlebnisses bei gleichzeitiger Steigerung des Eintauchens in das Spiel schließen [AC07, VKM05]. Dieses Eintauchen wird nach Csikszentmihalyi als Flow bezeichnet

¹ TU Braunschweig, Institut für Wirtschaftsinformatik, Abt. Informationsmanagement, Mühlenpfordtstr. 23, 38106 Braunschweig; linda.eckardt@tu-bs.de, a.pilak@tu-bs.de, m.loehr@tu-bs.de, p.van-treel@tu-bs.de, j.rau@tu-bs.de, s.robra.bissantz@tu-bs.de

[Cs90]. Dabei übt eine Person eine Tätigkeit konzentriert aus und geht vollkommen in dieser auf [Cs90]. Der im Rahmen eines Spiels auftretende Flow kann zu einem besseren Lernerfolg führen und ist somit ein Zustand, den die Entwickler von digitalen Game-based Learning Anwendungen erreichen wollen [Ho15]. Sweetser und Wyeth entwickelten eine GameFlow-Theorie, welche auf Basis der Flow-Theorie von Csikszentmihalyi zur Erklärung eines positiven Spielerlebnisses dient [SW05]. Ziel des Modells ist es, das Vergnügen von Spielen messbar zu machen. Ein Modell, welches auf diesen Arbeiten aufbaut, ist das EGameFlow-Modell (EGFM) von Fu et al. [FSY09]. Mithilfe dieses Modells wird das Spielerlebnis gemessen, während sich der Anwender beim Spielen in einer Game-based Learning Anwendung befindet [FSY09, Sh03, Ka12, Ho15]. Im Rahmen weiterer Forschung wurde die Skala des EGFM in den Arbeiten von Chu Yew Yee et al., Hoblitz, Vahldick et al. und Iten & Petko genutzt [CDQ10, Ho15, IP16]. Ein Ergebnis dieser Arbeiten ist, dass die Motivation bzw. das Spielerlebnis allein zu keinem signifikanten Lernerfolg führt. Vielmehr ist die Kombination aus dem Spielerlebnis und das Interesse an den Lerninhalten ein Indikator für hohen Lernerfolg [Ho15, IP16]. Vahldick et al. haben in ihrer Untersuchung eines Serious Games zum Lernen Erkenntnisse darüber erworben, welche Spielelemente verbessert werden müssen um ein besseres Spielerlebnis und folglich den Flow-Zustand zu erreichen bzw. zu halten [VMM15].

Im Vergleich zu den bisherigen Arbeiten, die das EGFM angewendet haben, wird in der vorliegenden Studie das vollständige EGFM mit allen Dimensionen (Konzentration, klare Zieldefinierung, Rückmeldung, Herausforderung, Autonomie, Immersion, soziale Eingebundenheit und Wissenszuwachs) anhand eines Serious Games zum Lernen von Informationskompetenz eingesetzt. Bisherige Arbeiten haben den Fokus auf einzelne Dimensionen gelegt. Beispielsweise verzichtete Vahldick et al. auf eine Prüfung der sozialen Eingebundenheit, da keine Spielelemente diese in der Anwendung förderten [VMM15]. Darüber hinaus untersuchten Hoblitz, Vahldick et al. und Iten & Pieko Serious Games mit anderen Lernthemen (z.B. programmieren lernen) [Ho15, VMM15, IP16]. Durch den Einsatz des vollständigen EGFM auf ein Serious Game zum Erlernen von Informationskompetenz sollen Erkenntnisse erlangt werden, inwiefern das Spielerlebnis ausgeprägt ist und was für Optimierungspotenziale vorhanden sind, um die Qualität des Serious Games in Hinblick auf den Lernerfolg zu verbessern.

2 Theoretische Grundlagen

2.1 EGameFlow

Aufbauend auf der Flow-Theorie von Csikszentmihályi und der GameFlow-Theorie von Sweetser und Wyeth, haben Fu et al. das EGFM entwickelt [FSY09]. Dieses befasst sich mit dem Flow bei EGames und stellt die These auf, dass ein Flow-Zustand in einem EGame zu einem besseren Lernerfolg führt und der auftretende Spielspaß das Lernen begünstigt. Fu et al. haben eine Skala zur Messung des EGameFlow entwickelt und im

Rahmen einer Studie validiert [FSY09].

Die Skala setzt sich in Anlehnung an Csikszentmihályi und Sweetser & Wyeth aus den acht Dimensionen Konzentration, klare Zieldefinierung, Rückmeldung, Herausforderung, Autonomie, Immersion, soziale Eingebundenheit und Wissenszuwachs zusammen [FSY09]. Um „Konzentration“ der Spielenden zu erhalten und zu steigern werden im Spiel Aktivitäten bereitgestellt, die eine Lernüberlastung minimieren. Je mehr Konzentration eine Aufgabe erfordert, desto mehr taucht der Spielende in das Spiel ein [FSY09, SW05]. Mit „Klare Zieldefinierung“ ist gemeint, dass die Spielenden klare Ziele rechtzeitig im Spiel aufgezeigt bekommen [Fe02]. Durch „Rückmeldung“ kennen die Spielenden jederzeit den aktuellen Spielstand und wissen, wie viel noch zur Fertigstellung einer Aufgabe fehlt. Unter „Herausforderung“ wird das Bereitstellen ebensolcher im Serious Game verstanden. Der Schwierigkeitsgrad der Herausforderung soll mit ansteigendem Level der Fähigkeiten der Spielenden variieren [FSY09, SW05]. Bei „Autonomie“ soll der Spielende die Initiative ergreifen und die vollständige Kontrolle über seine Entscheidung behalten [FSY09, Pa03]. Durch „Immersion“ sollen die Spielenden eine mühelose Beteiligung im Serious Game erleben. Dabei soll ein sorgenfreies und verändertes Zeitgefühl wahrgenommen werden. Unter „Soziale Eingebundenheit“ wird die Zusammenarbeit mit anderen Spielenden im Serious Game verstanden. Mit der Dimension „Wissenszuwachs“ ist ein Anstieg der Kenntnisse und Fähigkeiten der Spielenden gemeint [FSY09, SW05, Cs90].

2.2 Serious Game „Lost in Antarctica“

In der vorliegenden Studie wird das Serious Game „Lost in Antarctica“ (LIA), welches ein point-and-click Browserspiel ist, eingesetzt. Darin agieren die Spielenden innerhalb eines Forschungsteams in der Antarktis. Auf dem Weg zur Forschungsexpedition erleiden sie allerdings mit dem Flugzeug eine Bruchlandung, so dass neben der Forschung auch noch die Reparatur des Flugzeugs bewerkstelligt werden muss.

Die Spielenden erlernen in 12 Leveln unterschiedliche Fähigkeiten im Umgang mit Informationen (z.B. recherchieren oder wissenschaftliches Schreiben). Jedes Level ist identisch aufgebaut. Es findet eine Art der Wissensvermittlung (z.B. Video, Präsentation) statt und eine Wissensabfrage. Diese erfolgt über verschiedene Aufgabentypen, wie beispielsweise MC-Fragen, Lückentexte, Wortsuchen, Memory Spiele oder gemeinsam im Team zu lösende Aufgaben (z.B. Abstimmungen zu Fallbeispielen). Präsenzveranstaltungen werden als Teil eines Levels in Form von Forschungsmeetings punktuell eingesetzt, da einige Themengebiete eine intensivere Auseinandersetzung mit den Inhalten zum besseren Verständnis erfordern. In jedem Level können die Spielenden bis zu 300 Punkte erhalten, benötigen jedoch nur 200 Punkte, um im Serious Game voranzuschreiten. Für jeden erfolgreichen Abschluss eines Levels gibt es eine Belohnung (z.B. Bauteil für das Flugzeug oder Teamchat). Zusätzliche Punkte können auf einem Marktplatz gegen Minispiele, die ausschließlich der Unterhaltung dienen, eingetauscht werden. Darüber hinaus können sich die Spielenden gegenseitig bei der Reparatur

unterstützen, indem sie doppelte Bauteile auf einer Tauschbörse untereinander tauschen [ER16].

3 Studienbeschreibung

Das im Rahmen dieser empirischen Untersuchung herangezogene EGFM umfasst acht Dimensionen, die insgesamt aus 56 Items bestehen. Diese werden anhand einer 7-stufigen Likert-Skala (1 = trifft überhaupt nicht zu, ..., 7 = trifft voll und ganz zu) gemessen [FSY09, Ku13]. Für die Anwendung des EGFM ist zunächst eine Übersetzung der Items notwendig, da dieses Modell bislang nicht vollständig im deutschsprachigen Raum eingesetzt wurde [Ho15]. Um die Genauigkeit und Übereinstimmung der deutschen und englischen Versionen der Items zu gewährleisten wurde die Methode der Rückübersetzung gewählt [Ha96]. Nach einer Übersetzung der Autoren dieses Papers ins Deutsche wurde daher eine Rückübersetzung aller 56 Items durch eine Muttersprachlerin ins Englische durchgeführt. Damit sollte die Güte der Übersetzung festgestellt werden, um anschließend gegebenenfalls Nachbesserungen vorzunehmen. Einige Items wurden daraufhin angepasst, um die Übersetzung noch eindeutiger und inhaltlich genauer am englischen Original zu halten. Abbildung 1 zeigt die in der Umfrage verwendeten Items aller Dimensionen.

Studierende des Bachelorstudiengangs Wirtschaftsingenieurwesen Maschinenbau, die das Serious Game LIA im Rahmen ihres Studiums an der TU Braunschweig absolvieren müssen, haben im November 2016 an der Umfrage teilgenommen und den papierbasierten Fragebogen ausgefüllt. Insgesamt haben 83 Studierende die Fragebögen vollständig ausgefüllt. Mit 69 männlichen Teilnehmenden und 13 weiblichen Teilnehmenden haben überwiegend männliche Studierende an der Umfrage teilgenommen. Die Spannweite des Alters der befragten Studierenden reicht von 18 bis 31 Jahren bei einem Mittelwert von 22,21 Jahren. 84 % der Studierenden haben das Serious Game LIA zum Zeitpunkt der Umfrage bis zum vierten Level gespielt. Das entspricht ein Drittel des gesamten Serious Games. Die anderen befragten Studierenden haben weniger als vier Level gespielt.

4 Auswertung und Ergebnisse der Studie

Die Ergebnisse der abgefragten Items aus der Umfrage sind in Abbildung 1 dargestellt. Die allgemeine Zustimmung bzw. Ablehnung einzelner Items ist anhand der Mittelwerte (MW) erkennbar. Die Standardabweichung (SA) gibt die Streuung der Antworten an und ist ein Maß dafür, ob sich die Studierenden einig waren oder eine große Abweichung in den Antworten vorliegt.

Dimension	ID	ID und Frage	SA	MW
Konzentration	C1	Das Spiel ergreift meine Aufmerksamkeit	1,54	3,95
	C2	Das Spiel stellt einen Inhalt bereit, welcher meine Aufmerksamkeit stimuliert	1,42	3,46
	C3	Die meisten Spielaktivitäten beziehen sich auf das Lernen von Inhalten	1,43	4,98
	C4	Keine Ablenkung von der Aufgabe steht im Vordergrund	1,36	4,02
	C5	Im Großen und Ganzen kann ich die Konzentration im Spiel beibehalten	1,52	4,25
	C6	Bei Aufgaben, bei denen sich der Spieler konzentrieren soll, bin ich nicht abgelenkt	1,53	4,48
	C7	Ich würde nicht mit Aufgaben belastet, die in keinem Zusammenhang stehen	1,53	4,62
	C8	Der Arbeitsaufwand in diesem Spiel ist angemessen	1,65	4,44
Klare Zieldefiniertung	G1	Im Großen und Ganzen wurden die Ziele zu Beginn des Spiels aufgezeigt	1,57	4,44
	G2	Im Großen und Ganzen wurden die Ziele des Spiels klar aufgezeigt	1,46	4,63
	G3	Zwischenziele wurden zu Beginn jedes Kapitels aufgezeigt	1,17	4,91
	G4	Zwischenziele wurden klar und deutlich aufgezeigt	1,36	4,67
	G5	Ich verstehe die Lernziele durch das Spiel	1,45	4,87
Rückmeldung	F1	Ich erhalte Feedback zu meinem Spielefortschritt	1,41	5,02
	F2	Ich erhalte unmittelbares Feedback zu meinen Handlungen	1,34	5,00
	F3	Ich werde unmittelbar über neue Aufgaben benachrichtigt	1,30	4,63
	F4	Ich werde unmittelbar über neue Ereignisse benachrichtigt	1,31	4,85
	F5	Ich erhalte Informationen über meinen Erfolg (oder Misserfolg) von Zwischenzielen	1,07	5,45
	F6	Ich erhalte Informationen über meinen Status, wie etwa der Punktzahl oder des Levels	1,12	5,91
Herausforderung	H1	Ich genieße das Spiel, ohne Langeweile oder Angst zu verspüren	1,67	3,28
	H2	Die Herausforderung ist angemessen, weder zu schwierig noch zu leicht	1,34	4,85
	H3	Das Spiel stellt "Hinweise" im Text bereit, welche mir helfen, die Herausforderungen zu bewältigen	1,22	4,98
	H4	Das Spiel stellt eine "Online-Hilfe" bereit, welche mir hilft, die Herausforderungen zu bewältigen	1,27	4,33
	H5	Das Spiel stellt Videos oder Audiosequenzen bereit, welche mir helfen, die Herausforderungen zu bewältigen	1,05	5,72
	H6	Meine Fähigkeiten verbessern sich mit der Zeit durch die Bewältigung von Herausforderungen	1,42	4,46
	H7	Die Verbesserung meiner Fähigkeiten spornt mich an	1,50	3,94
	H8	Die Schwierigkeit der Herausforderungen erhöht sich, sobald sich meine Fähigkeiten verbessern	1,22	3,91
	H9	Das Spiel stellt in einem angemessenen Tempo neue Herausforderungen bereit	1,18	4,61
	H10	Das Spiel stellt verschiedene Schwierigkeitsstufen von Herausforderungen bereit, welche auf verschiedene Spieler zugeschnitten sind	1,41	3,78
	Autonomie	A1	Ich verspüre ein Gefühl der Kontrolle über das Menü (Starten, Abbrechen Speichern, etc.)	1,58
A2		Ich verspüre ein Gefühl der Kontrolle über die Aktionen von Rollen oder Objekten	1,56	4,50
A3		Ich verspüre ein Gefühl der Kontrolle über die Interaktion zwischen Rollen oder Objekten	1,53	4,38
A4		Das Spiel erlaubt es den Spielern nicht, Fehler in einem bestimmten Ausmaß zu machen, sodass die Spieler nicht mehr vorankommen können	1,91	3,85
A5		Das Spiel unterstützt mich dabei, aus Fehlern zu lernen	1,49	4,94
A6		Ich verspüre das Gefühl, dass ich meine Strategien frei wählen kann	1,72	3,74
A7		Ich verspüre ein Gefühl der Kontrolle und Einfluss auf das Spiel	1,71	3,66
A8		Ich kenne den nächsten Schritt im Spiel	1,45	3,64
A9		Ich verspüre ein Gefühl der Kontrolle über das Spiel	1,68	3,79
Immersion	I1	Ich vergesse jegliches Zeitgefühl, während ich das Spiel spiele	1,63	2,23
	I2	Ich nehme meine Umgebung nicht wahr, während ich das Spiel spiele	1,46	2,14
	I3	Ich vergesse zeitweise die Sorgen des alltäglichen Lebens, während ich das Spiel spiele	1,44	2,00
	I4	Ich erlebe ein verändertes Zeitgefühl	1,97	2,93
	I5	Ich kann mich im Spiel engagieren	1,52	3,23
	I6	Ich fühle mich emotional in das Spiel eingebunden	1,33	2,13
	I7	Ich fühle mich innerlich ins Spiel eingebunden	1,52	2,46
Soziale Eingebundenheit	S1	Ich fühle ein kooperatives Verhalten gegenüber anderen Mitspielern	1,41	2,93
	S2	Ich arbeite intensiv mit anderen Mitspielern zusammen	1,46	2,78
	S3	Die Kooperation in diesem Spiel ist für das Lernen hilfreich	1,54	3,27
	S4	Das Spiel unterstützt die soziale Interaktion zwischen Spielern (Chat, etc.)	1,59	3,58
	S5	Das Spiel fördert die Gemeinschaft innerhalb des Spiels	1,48	3,14
	S6	Das Spiel fördert die Gemeinschaft außerhalb des Spiels	1,61	2,91
Wissenszuwachs	K1	Das Spiel steigert mein Wissen	1,27	5,16
	K2	Ich verstehe den Grundgedanken des vermittelten Wissens	1,29	5,28
	K3	Ich versuche, das Wissen in diesem Spiel anzuwenden	1,32	5,02
	K4	Das Spiel motiviert den Spieler dazu, das gelehrte Wissen mit einzubeziehen	1,54	4,62
	K5	Ich will mehr über das gelehrte Wissen erfahren	1,71	3,83

Abb. 1: Ergebnisse des EGameFlow-Modells

Insgesamt zeigen die Ergebnisse der Dimensionen „Konzentration“, „klare Zieldefinierung“, „Rückmeldung“, „Herausforderung“, „Autonomie“ und „Wissenszuwachs“ eine positive Tendenz bei den Antworten der Studierenden, wohingegen die Dimensionen „Immersion“ und „soziale Eingebundenheit“ eine negative Tendenz aufweisen.

Die Dimension „Konzentration“ liegt im Vergleich zu den anderen Dimensionen im mittleren Feld. Sie hat eine leicht positive Tendenz, weist jedoch eine hohe Standardabweichung der Items mit Werten zwischen 1,36 und 1,65 auf. Eine Ursache dafür kann die Kombination aus Selbststudium (das Spielen des Serious Games) und Präsenzveranstaltung sein [ER16]. Laut Csikszentmihalyi sind Aufmerksamkeit und eine optimale Arbeitsbelastung unter anderem Indikatoren zur Konzentrationsförderung [Cs90]. Die Mischung aus Präsenzveranstaltung und Browserspiel könnte dies unterstützt haben. Um den positiven Trend weiter auszubauen, kann über eine detailliertere Spielwelt nachgedacht werden, da LIA bislang für alle Level ein identisches Design verwendet. Nach Johnson und Wiles können detailliertere Spielwelten jedoch ebenfalls Konzentration fördern [JW03].

Die Dimension „klare Zieldefinierung“ repräsentiert eine der drei am positivsten bewerteten Dimensionen. Dieser positive Trend kann damit begründet werden, dass das Hauptziel des Serious Games durch multimediale Unterstützung direkt am Anfang in Form eines Videointros präsentiert wird und ein frühzeitiges Aufzeigen der Ziele zur Förderung der Dimension beiträgt [Fe02, Pa03, ER16]. Federoff empfiehlt auch die Benutzung unterschiedlicher Aufgaben, die dem Spielenden deutlich gemacht werden [Fe02]. LIA setzt eine Checkliste in jedem Level ein, damit die Studierenden immer wissen, was zu erledigen ist und setzt vielfältige Aufgabentypen (z.B. MC-Fragen, Suchspiele, Kreuzworträtsel, Lückentexte) innerhalb der einzelnen Checklistenpunkte ein [ER16]. Gegebenenfalls könnte eine noch detailliertere Beschreibung der einzelnen Aufgaben innerhalb einzelner Checklisten dazu führen, dass die Dimension von zukünftigen Studierenden noch positiver wahrgenommen wird.

„Rückmeldung“ ist die Dimension mit den positivsten Bewertungen. Ein Vergleich mit den Ergebnissen von Fu et al. zeigt eine Übereinstimmung [FSY09]. Die Studierenden erhalten in LIA bereits Rückmeldung zu ihrem Fortschritt. In jeder Aufgabe der Checkliste wird angezeigt, was richtig oder falsch ist und wie viele Punkte noch notwendig sind, um ins nächste Level aufzusteigen. Das entspricht den Anforderungen an die Umsetzung der Dimension „Rückmeldung“ [Cs90, F02, La04, JW03, DCT04].

Die Dimension „Herausforderung“ ist im Vergleich mit den anderen Dimensionen durchschnittlich bewertet worden. Verglichen mit den Ergebnissen von Fu et al. wurde die Dimension des Serious Games in dieser Studie schlechter bewertet [FSY09]. Diese Information kann genutzt werden, um daraus Potentiale zur Verbesserung abzuleiten. Auffallend ist die Spannweite der Standardabweichung. Diese reicht in der Dimension von 1,05 bei Item H5 bis 1,67 bei H1. Hervorzuheben ist, dass der Einsatz von Multimedia innerhalb des Serious Games bei der Bewältigung der Aufgaben geholfen hat

Verbesserungspotential besteht hinsichtlich des Aufkommens von Langeweile oder Angst. Mit einem Mittelwert von 3,28 zeigt die Bewertung des zugehörigen Items eine negative Tendenz. Da die Schwierigkeit der Herausforderungen durchschnittliche Bewertungen erhalten hat, ist ein zu hoher oder zu niedriger Schwierigkeitsgrad für das Aufkommen von Langeweile oder Angst auszuschließen. Ein möglicher Grund für das Aufkommen von Langeweile in LIA ist, dass die Studierenden im Rahmen einer Pflichtveranstaltung dieses Serious Game spielen und nicht aus eigenem Interesse und somit nicht intrinsisch motiviert sind. An dieser Stelle greift die Relevanz des GameFlow, da diese die intrinsische Motivation fördert [BBV13]. Da in LIA auf eine Herausforderungsanpassung weitestgehend verzichtet worden ist [ER16], besteht in diesem Punkt Verbesserungsbedarf. Gegebenenfalls könnte ein Auflösen des statischen Spielablaufs mit den identisch aufgebauten Leveln und die Schaffung von zusätzlichen Handlungssträngen in der Spielgeschichte zu mehr Interesse an dem Serious Game führen und folglich das Aufkommen von Langeweile reduzieren [Ka12]. Darüber hinaus wird eine Anpassung des Schwierigkeitsgrads der Aufgaben empfohlen. Bislang sind unterschiedliche Schwierigkeitsstufen bei den Aufgaben innerhalb eines Levels integriert, jedoch aufgrund der unterschiedlichen Themengebiete nicht vollständig zwischen den Leveln [ER16]. Der Schwierigkeitsgrad sollte jedoch auch zwischen den Leveln einheitlich stufenweise anspruchsvoller werden, um das Interesse des Spielenden aufrechtzuerhalten, so dass dieser nach und nach sein Können steigert und somit die Herausforderungen bewältigen kann [DCT04, Pa03].

Das Ergebnis der Dimension „Autonomie“ liegt im Vergleich zu den anderen Dimensionen leicht unter dem Durchschnitt. Die Ergebnisse dieser Dimension sind in der Studie von Fu et al. im Vergleich besser ausgefallen [FSY09]. Es besteht also ein Verbesserungsbedarf in dieser Dimension. Die Mittelwerte liegen zwischen 3,64 bei Item A8 und 4,94 bei Item A5. Insgesamt liegt zusätzlich eine hohe Standardabweichung vor, die bei allen Items über 1,4 liegt und mit 1,91 bei Item A4 am höchsten ist. Verbesserungspotentiale ergeben sich bei den Items A4 und A6 bis A9 in Hinblick auf Kontrolle, Einfluss auf das Spiel und dem Gefühl den nächsten Schritt vorherzusehen. Positiv wird hingegen die Möglichkeit des Lernens aus Fehlern bewertet. Nach Sweetser und Johnson sollten für ein Gefühl der Kontrolle die Auswahl- und Handlungsmöglichkeiten in einem Spiel unbegrenzt sein [SJ04]. Dadurch kann der Spielende Spiele auf seine Weise spielen und Probleme auf seine Weise lösen. [SJ04] Bei LIA gibt es hingegen innerhalb der Level einen festen Spielablauf, bei dem einzelne Aufgaben in einer festen Reihenfolge gelöst werden müssen, da die Lerninhalte in dieser festen Reihenfolge gezeigt bzw. abgefragt werden. Somit kann zur Steigerung der Autonomie der Spielverlauf innerhalb der Level offener gestaltet werden, um den Spielenden größere Handlungsräume zu schaffen [ER16]. Dazu sollte in allen Leveln überprüft werden, ob eine festgelegte Reihenfolge der Aufgaben sinnvoll ist und ob in Teilen die Möglichkeit besteht, die Reihenfolge durch den Spielenden selbst festlegen zu lassen.

„Immersion“ ist die am schlechtesten bewertete Dimension dieser Umfrage. Auch im Vergleich zu den Ergebnissen von Fu et al. ist diese Dimension deutlich schlechter

bewertet worden [FSY09]. Am positivsten wird in Item I5 mit einem Mittelwert von 3,23 die Möglichkeit gesehen, sich im Serious Game engagieren zu können. Darüber hinaus gibt es auch in dieser Dimension eine hohe Standardabweichung, wobei diese bei Item I4 mit 1,97 den höchsten Wert der gesamten Umfrage aufweist. Insgesamt gibt es daher ein großes Verbesserungspotential in Bezug auf die Immersion. Allgemein führen der Einsatz von Erzählung, Sound (z.B. Soundeffekte oder Soundtracks) und einer guten Einführung in ein Spiel zu einem Gefühl der Immersion [SJ04, Sh08, RM04]. Bei LIA ist das Mittel der Erzählung stark ausgeprägt, da es einen durchgängigen Handlungsstrang gibt, an dem das gesamte Serious Game ausgerichtet ist. Eine stärkere Nutzung der Erzählung zur Steigerung der Immersion ist daher kaum möglich, jedoch eine Veränderung in der Art der Erzählung. Die Spielgeschichte wird bisher nämlich überwiegend in Textform mit Grafiken erzählt. Lediglich im Rahmen der Einführung in das Serious Game wird in einem Video auf die Spielgeschichte eingegangen. Zusätzlich zu diesem Einführungsvideo könnte die Spielgeschichte noch eindrucksvoller dargestellt werden, um die Einführung und somit die Immersion zu verbessern [SJ04]. Sound wird bislang nur in Videos zur Wissensvermittlung eingesetzt. Demzufolge kann der Einsatz von weiteren Soundelementen die Immersion im Serious Game erhöhen. Beispielsweise können während der Bearbeitung aller Level Musik gespielt oder beim falschen oder richtigen Beantworten einer Frage Soundeffekte eingesetzt werden [SJ04, RM04, Sh08].

Die Dimension der sozialen Eingebundenheit zeigt in dieser Umfrage eine negative Tendenz, wobei sich die Bewertungen der einzelnen Items zur Kooperation unter den Spielenden und zur Gemeinschaftsförderung nur unwesentlich voneinander unterscheiden und folglich auf einem ähnlichen Niveau sind. In den vier Serious Games, die Fu et al. untersucht haben, ist die Dimension der sozialen Eingebundenheit die am schlechtesten bewertete Dimension [FSY09]. Dies wird damit begründet, dass es technisch nicht leicht ist, ein interaktives Spiel für mehrere Spielende zu entwickeln [FSY09]. In dieser Studie ist die soziale Eingebundenheit zwar nicht am schlechtesten bewertet worden, jedoch am zweitschlechtesten. Ein möglicher Grund hierfür ist, dass die Studierenden, die an der Umfrage teilgenommen haben, maximal das vierte Level abgeschlossen haben und viele kooperative Spielelemente erst später im Spielablauf folgen. Für Spielende ist es jedoch wichtig miteinander zu kommunizieren, kooperieren und in Wettbewerb zu treten [La04]. In LIA wird allerdings erst mit Abschluss des dritten Levels der Teamchat freigeschaltet, so dass eine Kommunikation unter den Studierenden erst dann möglich ist [ER16]. Kooperative Aufgaben, kooperative Bewertungsverfahren und der Austausch von Bauteilen für die Reparatur des Flugzeugs folgen sogar erst in der Mitte des Serious Games [ER16]. Eine frühere Freischaltung von kooperativen Spielelementen ist empfehlenswert, um das Gefühl der sozialen Eingebundenheit zu erhöhen.

Die Dimension „Wissenszuwachs“ wurde insgesamt positiv bewertet. Im Vergleich zu den anderen Dimensionen variiert die Standardabweichung weniger. Bei den ersten drei Items (K1-K3) ist die Standardabweichung gering. Die Spielenden sind sich also einig, dass das Wissen mithilfe des Serious Games gesteigert wird (K1), der Grundgedanke des Wissens verstanden wird (K2) und das Wissen im Serious Game zudem angewendet wird (K3). Die Aussage von Item K5, in dem gefragt wird, ob Spielende mehr über das Gelernte

Wissen erfahren möchten, wird hingegen mit der höchsten Standardabweichung von 1,71 und einem Mittelwert von 3,83 am schlechtesten bewertet. Dies könnte an den Lerninhalten des Serious Games liegen, da die Vermittlung von Informationskompetenz mit einigen Herausforderungen verbunden ist. Beispielsweise gehen die Studierenden häufig davon aus, dass sie bereits alle notwendigen Kenntnisse im Umgang mit Informationen haben und daher eine Vertiefung des Wissens nicht unbedingt notwendig ist [MLR14].

Zusammenfassend sind in Abbildung 2 die Mittelwerte über alle Items jeder Dimension nochmal dargestellt.

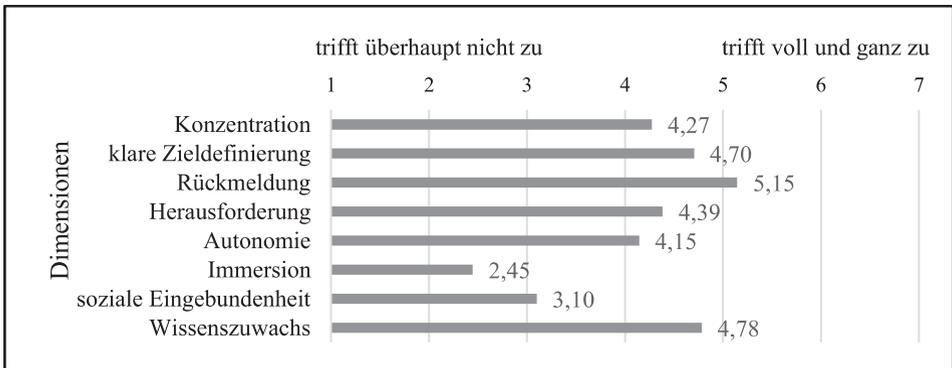


Abb. 2: Mittelwerte der Dimensionen

5 Schlussbemerkungen und Ausblick

In der vorliegenden Studie wurde das EGFM vollständig auf das Serious Game LIA angewendet. Während die Spielenden die Dimensionen „Rückmeldung“, „Wissenszuwachs“ und „klare Zieldefinierung“ positiv bewertet haben, konnten die Dimensionen „Konzentration“, „Herausforderung“ und „Autonomie“ nur Ergebnisse mit geringfügiger positiver Tendenz erzielen. Verbesserungspotential haben die Dimensionen „Immersion“ und „soziale Eingebundenheit“ ergeben. Ein verstärkter Einsatz von Videos zur Erzählung der Spielgeschichte und die Einbindung von Soundeffekten haben das Potential mehr Abwechslung zu generieren und das Gefühl der Immersion zu erhöhen. Die Freischaltung von kooperativen Spielelementen (z.B. Teamchat) zu Beginn des Serious Games ist ebenfalls empfehlenswert, um das Gefühl der sozialen Eingebundenheit zu steigern. Diese Empfehlungen sollten im Serious Game LIA umgesetzt werden, um den Lernerfolg der Studierenden positiv zu beeinflussen.

Um detailliertere Erkenntnisse zur Verbesserung des Spielerlebnisses und folglich des Lernerfolgs zu erhalten, ist in einer weiteren Untersuchung neben der quantitativen Messung eine qualitative Evaluation denkbar [St99]. Dabei haben die Studierenden die

Möglichkeit ihre Antworten zu begründen, wodurch konkretere Hinweise zur Verbesserung des Serious Games gegeben werden können.

Die Ergebnisse der Studie wurden durch den gewählten Befragungszeitpunkt beeinflusst, da die Studierenden zum Zeitpunkt der Befragung maximal das vierte Level erfolgreich abgeschlossen hatten. Insbesondere die kooperativen Elemente des Serious Games waren für die Studierenden dadurch noch nicht im vollen Umfang erkennbar, wodurch ein vollständiges Kennenlernen und Bewerten nicht möglich war. Es ist daher anzunehmen, dass die Ergebnisse bei der Wahl eines späteren Befragungszeitpunkts abgewichen wären.

Darauf aufbauend ist es empfehlenswert, die Befragung erneut durchzuführen, wenn die Studierenden das Serious Game vollständig durchgespielt haben. Dadurch kann eine Längsschnittstudie durchgeführt werden, die die Veränderung der Dimensionen im Verlauf des Serious Games aufzeigt und Rückschlüsse auf weitere Verbesserungspotentiale zulässt [Ri08].

Des Weiteren sollten die Empfehlungen zur Verbesserung des Serious Games integriert werden, um den Lernerfolg wirklich positiv beeinflussen zu können. Im Anschluss an die Integration sollte eine erneute Überprüfung des EGFM durchgeführt werden, um herauszufinden, ob die Studierenden wirklich eine Verbesserung der einzelnen Dimensionen wahrgenommen haben und so ein verbessertes Spielerlebnis empfunden haben. Neben der erneuten Überprüfung der einzelnen Dimensionen im Sinne eines iterativen Game-Design Prozesses [Fu14] sollten auch Einflussfaktoren des Lernerfolgs abgefragt werden (z.B. Interesse an Thematik, objektiver und subjektiver Wissensgewinn), um Erkenntnisse über die tatsächliche Beeinflussung der Dimensionen des EGFM auf den Lernerfolg treffen zu können. Bisher blieb dies aus und es konnten nur Implikationen aus dem EGFM zur Verbesserung des speziellen Serious Games in Hinblick auf die angestrebte Verbesserung des Lernerfolgs getroffen werden.

Literaturverzeichnis

- [AC07] Ampatzoglou, A.; Chatzigeorgiou, A.: Evaluation of object-oriented design patterns in game development. *Information and Software Technology* 49/5, S. 445–454, 2007.
- [BBV13] Balogh, A.; Benedek, A.; Vidékiné-Reményi, J.: Who is the Joy of Learning Important for? In (Flate Paulsen, M.; Szücs, A., Hrsg.): *Proceedings of The joy of learning: Enhancing learning experience improving learning quality - EDEN 2013 conference*, Oslo, 2013.
- [CDQ10] Chu Yew Yee, S. L.; Duh, H. B. L.; Quek, F.: Investigating narrative in mobile games for seniors. In (Mynatt, E. et al., Hrsg.): *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, S. 669-672, 2010.
- [Cs90] Csikszentmihalyi, M.: *Flow: The psychology of optimal experience*, Harper Perennial, New York, 1990.
- [DCT04] Desurvire, H.; Caplan, M.; Toth, J.A.: Using heuristics to evaluate the playability of

- games. In (Dykstra-Erickson, E.; Tscheligi, M., Hrsg.): Extended Abstracts of the 2004 Conference on Human Factors in Computing Systems, Vienna, S. 1509-1512, 2004.
- [De07] Deterding, S. et.al.: Gamification. Toward a Definition. Paper presented at the CHI Workshop Gamification, VanCouver, Canada, 2007.
- [ER16] Eckardt, L.; Robra-Bissantz, S.: Design eines Spiels zum Lernen von Informationskompetenz. In (Lucke, U.; Schwill, A.; Zender, R., Hrsg.): DeLFI 2016 - 14. E-Learning Fachtagung Informatik, Lecture Notes in Informatics (LNI), Potsdam 2016, S. 15-26, 2016.
- [Fe02] Federoff, M.: Heuristics and usability guidelines for the creation and evaluation of fun in video games 2002, http://ocw.metu.edu.tr/file.php/85/ceit706_2/10/MelissaFederoff_Heuiristics.pdf, Stand: 14.01.2016.
- [Fu14] Fullerton, T.: Game design workshop: a playcentric approach to creating innovative games. CRC Press, 2014.
- [FSY09] Fu, F.-L.; Su, R.-C.; Yu, S.-C.: EGameFlow: a scale to measure learners' enjoyment of e-learning Games. *Computers & Education* 52/1, S. 101–112, 2009.
- [Ha96] Hambleton, R. K.: Guidelines for Adapting Educational and Psychological Tests. National Center for Education Statistics, S. 14-15, 1996.
- [Ho15] Hoblitz, A.: Spielend Lernen im Flow: Die motivationale Wirkung von Serious Games im Schulunterricht, Springer, Wiesbaden, 2015.
- [IP16] Iten, N.; Petko, D.: Learning with serious games: Is fun playing the game a predictor of learning success? *British Journal of Educational Technology* 47/1, S. 151–163, 2016.
- [JW03] Johnson, D.; Wiles, J.: Effective affective user interface design in games. *Ergonomics* 16/13-14, S. 1332-1345, 2003.
- [Ka12] Kapp, K. M.: The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education, Pfeiffer, San Francisco, 2012.
- [Ku13] Kuckartz, U. et al.: Statistik: eine verständliche Einführung, Springer, Wiesbaden, 2013.
- [La04] Lazzaro, N.: Why We Play Games: Four Keys to More Emotion Without Story, http://www.xeodesign.com/whyweplaygames/xeodesign_whyweplaygames.pdf, Stand: 28.12.16.
- [MLR14] Markey, K.; Leeder, C.; Rieh, S. Y.: Designing Online Information Literacy Games Students Want to Play. Rowman & Littlefield, Maryland, 2014.
- [Pa03] Pagulayan, R. et al.: The Human-Computer Interaction Handbook: Fundamentals, Evolving Techniques and Emerging Applications. User-centered design in games, Lawrence Erlbaum Associates, New York, 2003.
- [Pr01] Prensky, M.: Digital game-based learning. Types of learning and possible game styles, McGraw-Hill, 2001.
- [Ri08] Rindfleisch, A. et al.: Cross-Sectional Versus Longitudinal Survey Research: Concepts, Findings and Guidelines. *Journal of Marketing Research* 45/3, S. 261-279, 2008.
- [RA03] Rollings, A.; Adams, E.: On game design, New Riders, Indianapolis, 2003.

- [RM04] Rollings, A.; Morris, D.: *Game Architecture and Design: A New Edition*, New Riders, Indianapolis, 2004.
- [Sh08] Shell, J.: *The Art of Game Design. A Book of Lenses*, Morgan Kaufmann Publishers, Burlington, 2008.
- [Sh03] Shernoff, D. J. et al.: Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly* 18/2, S. 158-176, 2003.
- [SJ04] Sweetser, P.; Johnson, D.: Player-centred game environments: Assessing player opinions, experiences and issues. In (Rauterberg, M., Hrsg.): *Proceedings Entertainment Computing - ICEC 2004: Third International Conference*, Eindhoven, S. 321-332, 2004.
- [St99] Stier, W.: *Empirische Forschungsmethoden*, Springer Verlag, Berlin u.a., 1999.
- [SW05] Sweetser, P.; Wyeth, P.: GameFlow: A model for evaluating player enjoyment in games. *ACM Computer in Entertainment* 3/3, S. 1–24, 2005.
- [VMM15] Vahldick, A.; Mendes, A. J.; Marcelino, M. J.: Analyzing the Enjoyment of a Serious Game for Programming Learning With two Unrelated Higher Education Audiences. In (Munkvold, R., Kolas, L. Hrsg.): *European Conference on Games Based Learning*, Steinkjier, S. 523-531, 2015.
- [VKM05] Virou, M.; Katsionis, G.; Manos, K.: Combining software games with education: evaluation of its educational effectiveness. *Educational Technology and Society* 8/2, S. 54–65, 2005.

Spielend leicht Lernspiele entwickeln - Ein Framework für Multitouch-Lernspiele

Matthias Ehlenz¹, Thiemo Leonhardt¹ und Ulrik Schroeder¹

Abstract: Kollaborative Lernspiele können einen entscheidenden Beitrag zum gemeinschaftlichen Lernen in einem Klassenzimmer der Zukunft leisten. Dieser Betrag stellt die Arbeit an einem Framework vor, das künftig zum einen die Entwicklung von Lernspielen für Multitouch-Geräte aller Größen und Plattformen erleichtert, zum anderen deren Struktur vereinheitlicht und so eine systematisierte Beforschung ermöglicht. Es werden Struktur und Ziele des Projektes vorgestellt, der Aufbau des Frameworks und seiner Kernkomponenten erörtert und schließlich der aktuelle Stand und die geplante Weiterentwicklung dargelegt.

Keywords: Multitouch, Lernspiele, Gamification, Framework-Entwicklung, Learning Analytics, adaptives Feedback, Kollaboration, Multiuser

1 Einleitung

Elektronisch unterstütztes Lernen bietet im Zeitalter globaler Vernetzung nie dagewesene Chancen und Perspektiven. Aktuelle Entwicklungen bringen digitale Lernmedien innerhalb von Bildungseinrichtungen und darüber hinaus mit Lernenden aller Altersgruppen in Kontakt. Insbesondere junge Lernende sind mit digitalen Medien aufgewachsen und dadurch bereits einen hohen (grafischen) Standard gewöhnt. Diese Herausforderung muss von spielerischen Lernformen wie im Themenfeld gemeistert werden, Gamification ist das Stichwort.

Darüber hinaus bieten digitale Lernmedien auf vernetzten Geräten die Chance individuell auf die Lernenden einzugehen, das Lernverhalten zu analysieren und den Lernprozess so individualisiert zu unterstützen. Während die technischen Voraussetzungen hierfür heute als gegeben betrachtet werden können, so erfordert eine entsprechende Umsetzung tiefgehende Fachkenntnisse auf vielen Gebieten. Neben dem domänenspezifischen Fachwissen, welches vermittelt werden soll, und den fachdidaktischen Kenntnissen, die Methodik und Aufbereitung der Inhalte einbringen, sind Erfahrungen in der Gestaltung von Benutzeroberflächen, Kenntnisse in der Programmierung von Spielen und der Einsatz von Learning Analytics für Feedback und Erforschung der Lernprozesse erforderlich. Dieser Beitrag erläutert die laufende Arbeit diesen Prozess zu systematisieren und automatisiert zu unterstützen.

¹ RWTH Aachen, Learning Technologies Research Group, Ahornstraße 55, 52074 Aachen, nachname@cs.rwth-aachen.de

2 Struktur und Ziele

Ziel der Forschungsgruppe ist die Implementation eines Multitouch-Learning-Game-Framework (MTLG), das sowohl wissenschaftliche als auch unterrichtspraktische Zwecke erfüllt. Auf forschender Seite wird die Untersuchung gezielten Feedbacks zur Unterstützung und Förderung kollaborativer und kooperativer Gruppenprozesse im gemeinschaftlichen Lernen an großen Multitouch-Displays ermöglicht. Aus Perspektive von Lehrenden unterstützt das Framework die Entwicklung von Lernspielen als Medien für den Unterricht. Aufgrund der Angliederung der Informatik-Lehramtsausbildung an der Forschungsgruppe sind die Umsetzungen der Lernspiele bisher alle im Bereich der Informatik-Didaktik erfolgt.

Didaktisches Ziel des Projektes ist es, kollaborative, fokussierte und kurzweilige elektronische Lernspiele zu erstellen, die Unterrichtsinhalte durch gezielte motivationssteigernde Spielelemente aufwerten. Dabei wird vor einer freien Spielmöglichkeit ein spielspezifisches Scaffolding vorangestellt [HP97]. Aus technischer Sicht setzt das Projekt ausschließlich auf Technologien, die eine möglichst vollständige Unabhängigkeit von Endgeräten und Betriebssystemen garantieren.

Touchscreens sind ein Medium, das für alle Altersgruppen von Lernenden spezifische Vorteile mit sich bringt und eine intuitive Nutzung ermöglicht [Pe14] [FKR13]. Neben dem Einsatz der Lernspiele auf Tablets konzentriert sich das Projekt als weiteres Einsatzszenario auf Multitouchtische. Neben mittleren Devices (27 Zoll) mit bis zu 30 simultanen Berührungseingaben für bis zu vier gleichzeitige Spieler werden die Spiele auch auf neuesten Devices wie dem Microsoft Hub (84 Zoll) mit 100 gleichzeitigen Berührungen umgesetzt. Geräte mit der Möglichkeit, mit mehreren Lernenden gleichzeitig kollaborativ an einem Spiel zu partizipieren, sehen wir als einen wesentlichen Bestandteil des Klassenzimmers der Zukunft.

2.1 MTLG als Forschungsinstrument

Die didaktische und technische Grundlage zur Software-Architektur des Frameworks lieferten Vorstudien in den Bereichen Gamification [HSS15] und Feedback in Multitouchlernspielen [Sc13] [Ho12], [Na11]. Weiterhin fließen praktische Erfahrungen aus einem regelmäßigen Software-Projektpraktikum im Fachbereich Informatik an der RWTH Aachen mit Studierenden aus den Bachelor Studiengang Informatik sowie aus dem Lehramtsstudiengang Informatik in die Weiterentwicklung des Frameworks mit ein [EBS16]. Die Auswahl der Lerninhalte sowie deren didaktische Reduktion und Anpassung sind der Fachdidaktik Informatik zugeordnet, da hier mit dem Informatik-Schülerlabor InfoSphere ein passendes Testbed vorhanden ist. Die didaktisch entwickelten Spielabläufe, Unterstützungen und Rückmeldungen an die Lerner sind aber auf vergleichbare fachübergreifende Lerninhalte übertragbar.

2.2 Umgesetzte Spiele

Bisher wurden insgesamt zehn Lernspiele fertiggestellt. Die Themenbereiche bewegen sich im gesamten Spektrum der Informatik und reichen von Spielen für die Grundschule bis hin zur gymnasialen Oberstufe (hier in Abstimmung mit dem Lehrplan der Sek. II/NRW). Die Spiele thematisieren beispielsweise Sortieralgorithmen, Datenstrukturen, reguläre Ausdrücke, boolesche Algebra und strukturierte Zerlegung. Die Erprobung erfolgt im Schülerlabor InfoSphere mit Schülerinnen und Schülern der intendierten Zielgruppen. Eine ausführlichere Darstellung der Inhalte erfolgte in [LES17].

2.3 Die Software-Architektur

Bei der Entwicklung des Frameworks stehen aus technischer Sicht Zukunftstauglichkeit, Offenheit und Kompatibilität im Vordergrund. Aus diesen Gründen wurden als Basistechnologien HTML5 und JavaScript gewählt. HTML5 führt das neue Canvas-Element ein, welches vollständige Flexibilität und gestalterische Freiheit mit sich bringt, JavaScript hat sich von einer Außenseitersprache für Webentwickler zu einem allgemeinen Werkzeug entwickelt, das als interpretierte Skriptsprache dem Entwickler vollständige Transparenz bietet.

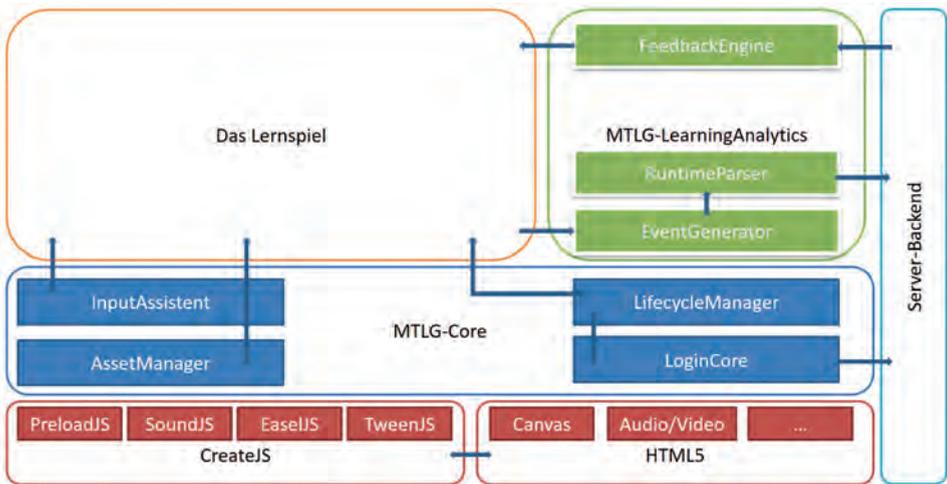


Abb. 1: (geplanter) Aufbau des Frameworks

Um den Umgang mit dem Canvas-Element zu erleichtern und nicht grundlegende Zeichenfunktionen neu implementieren zu müssen, wurde hier auf ein Open-Source Bibliotheken-Set CreateJS zurückgegriffen, das neben den Zeichen- und Animationsbibliotheken EaselJS und TweenJS mit PreloadJS und SoundJS weitere Möglichkeiten zum Vorladen von Assets und dem Abspielen von Tönen mit sich bringt. CreateJS hat sich in Vortests in Offenheit, Kompatibilität und Performance gegen die gegenwärtig

erhältlichen Alternativen durchgesetzt.

Kern des Frameworks sind Komponenten (siehe Abb. 1), die das Zusammenspiel der Module sicherstellen und dem Entwickler eine zusätzliche Abstraktionsebene bieten und so eine strukturierte Spieleentwicklung unterstützen. Dazu gehört die Unterstützung im Umgang mit Bildern und Tönen (Verwaltung der grafischen Objekte insbesondere des Vorladens und des Caching, um Performanceproblemen vorzubeugen), die einfache Einbindung vielfältiger Eingabegeräte über konventionelles Multitouch hinaus sowie die Entwicklung komplexer (auch nichtlineare) Levelverläufe durch ein Lifecycle-Management. Zudem abstrahiert das Framework das Usermanagement und stellt Funktionalitäten zur kindgerechten Nutzerverwaltung und -identifikation zur Verfügung, mit denen sich Entwicklerinnen und Entwickler so nicht weiter auseinandersetzen müssen.

Insbesondere für Kooperation und Kollaboration bietet das Input-Modul Unterstützung für spezielle Eingabegeräte, die mit kapazitiven Multitouch-Displays interagieren können, sogenannte Tangibles. Diese erweitern den gleichberechtigten Workspace des horizontalen Displays um haptische Elemente, die Lerninhalte „begreifbar“ machen können. Die Entwicklung für Tangibles ist mit Hilfe des Frameworks nicht aufwändiger als die Verarbeitung konventioneller Toucheingaben.

2.4 Learning Analytics als Forschungsvertiefung

Das MTLG-Framework vereinheitlicht die Entwicklung von Lernspielen und stellt Schnittstellen zu einem Learning-Analytics-Modul zur Verfügung. Das Learning-Analytics-Modul besteht derzeit aus drei Unterkomponenten: Der Eventgenerator erleichtert es Entwicklern standardkonforme Event-Objekte zur späteren Analyse zu generieren. Der Runtime-Parser ermöglicht es dem Entwickler eigene Event-Sequenzen durch Pattern zu definieren, die zur Laufzeit erkannt werden und Meta-Events triggern und Callback-Funktionen aufrufen. Die (noch nicht eingebundene) Feedback-Engine ermöglicht es dem Spiel, dynamisch mit Feedback auf Spielsituationen zu reagieren. Die Vereinheitlichung der Spielentwicklung ermöglicht es, Erweiterungen der Learning Analytics Module in die schon implementierten Spiele zu integrieren. Durch den simultanen Einsatz dieser zentralen Erfassung und konventionelle qualitative Beobachtung können Rückschlüsse auf den Einzelnen sowie die Wirksamkeit auf das Kollaborationsverhalten in der Gruppe gezogen werden.

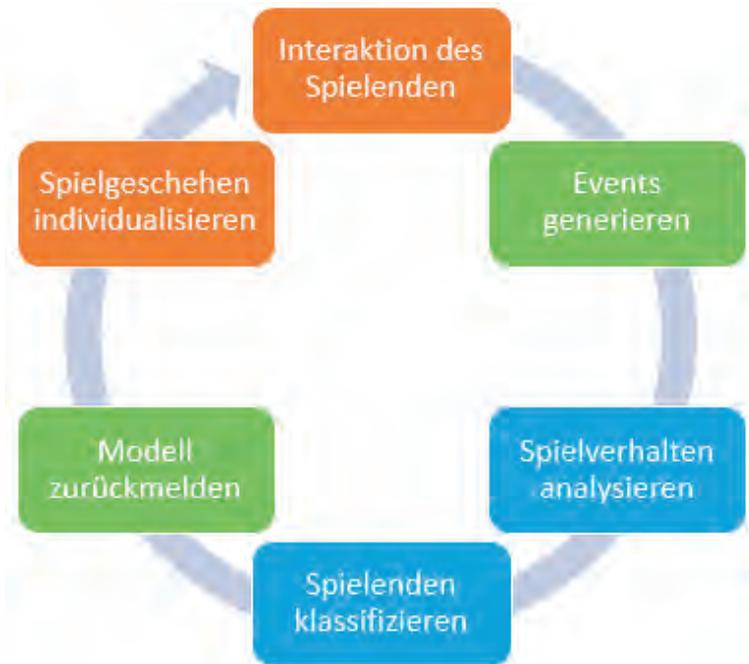


Abb. 2: Prozess der Erfahrungsindividualisierung

Die Learning Analytics-Komponente des Frameworks wird, ebenso wie der Spielverlauf, vom Lifecycle-Manager kontrolliert. Der Entwickler des Spiels legt hierbei fest, zu welchem Zeitpunkt im Spiel Events geworfen werden und ordnet diese sowohl Spielern als auch vordefinierten Kategorien zu. Zusätzlich besteht für den Entwickler die Möglichkeit, über Entities in der Objektstruktur zusätzliche Informationen in Form von Key/Value-Paaren zuzuordnen, um sowohl eine spielspezifische als auch generische Auswertung zu ermöglichen.

3 Ausblick und Forschungsperspektive

Neben der praxisorientierten Weiterentwicklung des Frameworks ist dieses primär als Forschungsinstrument konzipiert. Der Schwerpunkt der avisierten Forschung liegt hierbei zunächst auf adaptiven Feedback-Komponenten, die kooperatives und kollaboratives Lernen fördern, sowie den Learning-Analytics-Bereichen, die neben diesen simultanen Rückmeldungen den Lernenden nachgelagert, individualisiert und privat eine Beobachtung des Lernfortschritts ermöglichen können. Darüber hinaus ermöglichen die LA-Komponenten Lehrkräften eine strukturiertere Reflexion der Erreichung der Unterrichtsziele, eine Perspektive, die weitere Untersuchungen

legitimiert. Hier anknüpfend ist auch die inhaltliche Beforschung der Spiele möglich. Diese fachdidaktische Sicht befasst sich mit der Spieleforschung, z.B. deren Passung zum Lehrplan und die Eignung zur Vermittlung und/oder Überprüfung der Lehrinhalte.

Als Ausblick wird das Framework und die entwickelten Spiele langfristig im Sinne der Open-Educational-Resources öffentlich gemacht und ermöglicht somit nicht nur einer breiteren Basis von Entwicklern die Mitwirkung, sondern darüber hinaus interessierten Lehrkräften die Entwicklung eigener Lernspiele und der Didaktik weiterer Domänen die Entwicklung angepasster Lernspiele für fachspezifische Inhalte.

Literaturverzeichnis

- [EBS16] Ehlenz, M.; Bergner, N.; Schroeder, U.: Synergieeffekte zwischen Fach- und Lehramtsstudierenden in Softwarepraktika. *Hochschuldidaktik der Informatik*, S. 99, 2016.
- [FKR13] Feierabend, S.; Karg, U.; Rathgeb, T.: KIM-Studie 2014. In *Kinder+ Medien, Computer+ Internet. Basisuntersuchung zum Medienumgang 6-bis 13-Jähriger in Deutschland*, 2013.
- [Ho12] Holz, J. et al.: Serious Games on Multi Touch Tables for Computer Science Students: CSEDU (2). *SciTePress*, S. 519–524, 2012.
- [HP97] Hogan, K.; Pressley, M., Hrsg.: Scaffolding student learning. *Instructional approaches and issues*. Brookline Books, Cambridge, MA, 1997.
- [HSS15] Hurtienne, D.; Schroeder, U.; Spannagel, C.: IT EnGAGES!–Adaptierbare Gamification in einer Anfänger-Programmervorlesung. In *HDI 2014–Gestalten von Übergängen: 6. Fachtagung Hochschuldidaktik der Informatik*; 15.-16. September 2014, Universität Freiburg 9, S. 27, 2015.
- [LES17] Leonhardt, T.; Ehlenz, M.; Schroeder, U.: Kollaborative Multiuser Serious Games für elektronische Schulbücher. In (Schuhen, M.; Froitzheim, M. Hrsg.): *Das elektronische Schulbuch. Fachdidaktische Anforderungen und Ideen treffen auf Lösungsvorschläge der Informatik*. Lit-Verl., Berlin, 2017.
- [Na11] Nabbi, N. et al.: Kollaborative und altersgerechte Lernanwendung zur Vermittlung fundamentaler Ideen der Informatik. In *DeLFI 2011–Die 9. e-Learning Fachtagung Informatik*, 2011.
- [Pe14] Peetz, G.: Mit den Fingern die Welt erkunden. Ein Forschungsprogramm zur motorischen und haptischen Nutzung des Touchscreens aus Sicht der Kleinkind- und Kinderzeichnungsforschung. In *Diskurs Kindheits- und Jugendforschung* 9, S. 319–338, 2014.
- [Sc13] Schäfer, A. et al.: From boring to scoring – a collaborative serious game for learning and practicing mathematical logic for computer science education. *Computer Science Education* 23, S. 87–111, 2013.

Simulationsfall Nohra: SimCity als etabliertes Lehrmedium in der universitären Hochschulausbildung

Uwe Arnold^{1,2}, Heinrich Söbke², Maria Reichelt³ und Thomas Haupt²

Abstract: Simulationsspiele bieten einen geschützten Raum, um dort mit modellierten Systemen zu experimentieren. Sie ermöglichen Lernerfahrungen zu wechselseitigen dynamischen Abhängigkeiten und fördern damit den Aufbau eines tiefergehenden Systemverständnisses, der sich durch die aktivere kognitive Informationsverarbeitung auch förderlich auf den Lernerfolg auswirken kann. Diese Eigenschaften haben zu einer vielfältigen Verwendung von Simulationsspielen in der Lehre geführt. Wir stellen hier den Einsatz von SimCity in der Vorlesung Infrastrukturmanagement vor, der seit 2002 kontinuierlich betrieben wird. Damit liegen langjährige Erfahrungen eines Simulationsspiels als Lernwerkzeug vor. Wir beschreiben den spezifischen Aufbau dieses Einsatzes und mögliche erfolgskritische Faktoren. Methodisch greifen wir dabei u.a. auf Veranstaltungsbeobachtungen und qualitative Befragungen zurück. Weiter gehen wir auch auf die Herausforderungen ein, die insbesondere aus der langen Einsatzzeit resultieren (*Game-Aging*). Insgesamt belegt dieser Einsatz, dass Simulationsspiele zu einem etablierten Lernmedium gereift sind.

Keywords: SimCity; Simulationsspiel; Fallstudie; Game-Aging; Instructional Design;

1 Simulationsspiele als Lernwerkzeug

Computerspiele werden zu den *Neuen Medien* gezählt und wurden schon früh als Lernwerkzeuge identifiziert [Sq03]. Auch aus Sicht des *Instructional Designs* sind Spiele geeignete Lehrmittel, wenn sie didaktisch sinnvoll eingesetzt werden [Ma16]. Dazu müssen sie motivations- und lernförderlich gestaltet sein, indem sie auch die kognitive Belastung des Spielenden gering halten (u.a. [BPM10],[Sw06]). Simulationsspiele eignen sich insbesondere als Experimentierumgebungen für den Umgang mit komplexen Systemen [Dö81]. Der Einsatz von SimCity als das wohl bekannteste und kommerziell erfolgreichste Städtebauspiel wurde in vielen unterschiedlichen Kontexten bereits untersucht. Bisher liegen jedoch nur wenige empirische Arbeiten vor, in denen der Einsatz von SimCity über einen längeren Zeitraum dokumentiert wurde ([Ga07], [Wo15]).

2 Nohra als SimCity-Spielszenario

Der Kurs *Urbanes Infrastrukturmanagement* ist ein Bestandteil des Master-Studiengangs

¹ AHP GmbH & Co. KG, Karl-Heinrich-Ulrichs-Strasse 11, 10787 Berlin, arnold@ahpkg.de

² Bauhaus-Universität Weimar, Bauhaus-Institut für zukunftsfähige Infrastruktursysteme, Coudraystr. 7, 99423 Weimar, {heinrich.sobke|thomas.haupt}@uni-weimar.de

³ Fachhochschule Erfurt, Zentrum für Qualität, PF 45 01 55, 99051 Erfurt, maria.reichelt@fh-erfurt.de

Umweltingenieurwesen an der Bauhaus-Universität Weimar. In einem vorbereitenden Projekt [HAB05] hatte sich herausgestellt, dass zunächst SimCity 3000 und später SimCity 4 geeignet waren, wesentliche Problemstellungen der regionalen Infrastrukturentwicklung zu illustrieren. Die Nutzung dieses Werkzeuges in einer Fallstudie profitiert dabei von den Erfahrungen, die der Erstautor als Projektmanager der Infrastrukturentwicklung der Gemeinde Nohra und des Industriegebietes UNO (Ulla, Nohra, Obergrundstedt) in der Nähe von Weimar in Thüringen machen konnte. Daher wurde im Wintersemester 2002 zum ersten Mal das Simulationsspiel SimCity 4 unter Nutzung des Szenarios Nohra eingesetzt. Bemerkenswert an diesem Szenario ist zum einen die modellhafte Nachbildung eines tatsächlichen Ausgangsstandortes nach der Wende mit geringer Entfernung zur Bauhaus-Universität Weimar. Zum anderen ist die tatsächliche Entwicklung des Standorts bekannt. So ist ein Vergleich der Simulationsergebnisse mit der realen dynamischen Entwicklung möglich. Im Anschluss an die Simulationen wird regelmäßig eine Exkursion der Studierenden zum Standort Nohra durchgeführt.

Zur Erstellung eines Szenarios in SimCity wird zunächst durch die Festlegung von Wohn-, Industrie- und Gewerbegebieten eine Art Bauleitplanung durchgeführt. Auf Basis dieser Planung werden dann die angebotenen Modellelemente, wie Straßen, Wasser- und Energieversorgungsinfrastruktur und spezielle Bauwerke (z.B.: Abwasserbeseitigungsanlage) derart platziert, dass sie das angestrebte Zielmodell – wie in diesem Fall Nohra – möglichst realitätsnah nachbilden. Dabei gilt jedoch, dass die quantitative Abbildung der simulierten Teilprozesse teilweise nur sehr eingeschränkt gelingt. Dies betrifft sowohl die geographische Ähnlichkeit, als auch die sozio-ökonomische Ähnlichkeit (flächenspezifische demographische und Beschäftigungsentwicklung, spezifische Investitionsaufwände, Betriebskosten, ...). Anschaulichkeit und Benutzerfreundlichkeit des interaktiv-graphischen Simulationssystems sind dagegen sehr hoch.

Urbanes Infrastrukturmanagement wird als semesterfüllende Blockveranstaltung mit jeweils 3 Doppel-Unterrichtsstunden pro Block-Tag durchgeführt. Der Einsatz von SimCity erfolgt an einem dieser Tage. Dieser ist wie folgt gegliedert: Zunächst wird eine ca. 90-minütige Einführung zum konkreten Anwendungsfall Nohra und seiner realen zeitlichen Entwicklung gegeben. Darauf folgt die Spielphase. Jede Gruppe hat einen eigenen Rechner mit angeschlossenem Beamer zur Verfügung. Die Gruppengrößen variieren – abhängig von der Beteiligung – von 2 bis 7 Personen. Die Spielphase selbst ist in drei separate Einzelphasen von je ca. 30 Minuten Länge unterteilt: Der erste Abschnitt wird genutzt, um Vertrautheit mit der Spielsoftware herzustellen (ohne Wertung der Ergebnisse). Die Aufgabe des zweiten Abschnitts ist es, bei gegebener Anfangs-Bauleitplanung das Szenario möglichst gut zu entwickeln, im dritten Abschnitt darf die Bauleitplanung selbständig vorgegeben werden. Nach jedem Szenario erfolgt eine kursgemeinsame Bewertung der erreichten Spielstände unter Nutzung eines Tabellenkalkulationswerkzeuges. Es wird die Rangfolge der Gruppen festgestellt. Während des Spiels werden Hinweise gegeben, die gemeinsam mit den zu Beginn definierten Kriteriengewichten als Grundlage der Spielentscheidungen dienen. In einer Nachbesprechung werden die wesentlichen Spiel-Erfahrungen diskutiert.

3 Untersuchungsziele und Methodik

Das kennzeichnende Merkmal dieses Einsatzes von SimCity in einem formalen Bildungskontext ist der lange Erfahrungshorizont von nunmehr fast 15 Jahren. Ein Ziel der Studie war es, zu untersuchen, was einen erfolgreichen Einsatz eines Computerspiels in der Hochschullehre kennzeichnet. Darüber hinaus sollen die Akzeptanz und subjektiven Einschätzungen der Studierenden bei der Anwendung dieses Lehr-Lernkonzepts analysiert werden. Zur Untersuchung wurden zwei Veranstaltungen beobachtet (WS 2015/16 mit $n=19$ und WS 2016/17 mit $n=7$ Teilnehmenden) und ein leitfadengestütztes Interview mit der Lehrperson durchgeführt. Die Teilnehmenden der zweiten Veranstaltung wurden mit Hilfe eines Fragebogens (ca. 90 Items) befragt und nahmen an einer strukturierten, ca. 30 min dauernden Gruppendiskussion teil.

4 Ergebnisse

4.1 Kennzeichen des Spiel-Szenarios

Wichtiges Merkmal ist die offene Spielweise, d.h. zu jedem Zeitpunkt können alle Teilnehmer den Spielstand der jeweils anderen Gruppen beobachten und davon lernen. Zum Beispiel zeigte sich das Phänomen, dass aktiv „Kundschafter“ ausgesandt wurden, um die anderen Gruppen „zu beobachten“. Dieses war zuvor schon durch Hinweise des Dozierenden angeregt worden. Dieser forderte außerdem ein, dass sich die Gruppen einen Namen geben sollten. Bei Mangel an kreativen Ideen konnte er genauso aushelfen (beispielsweise entstand so die Gruppe *TFB – The Future is bright*). Die einfallsreichen Gruppennamen trugen nicht nur zu einer erhöhten Gruppenidentifikation, sondern auch zu einer entspannten Lernatmosphäre bei. Ebenfalls wurde beobachtet, dass eine exzellente Kenntnis der Spielsoftware sehr hilfreich ist, um die Studierenden in der relativ kurzen Zeit von 30 Minuten in die Lage zu versetzen, zielgerichtete Aktionen zur Entwicklung des Szenarios durchzuführen. Ein fortlaufender Rundgang durch die Gruppen mit Hinweisen und aktivierenden Kommentaren sorgte für eine lockere, aber dennoch fokussierte Stimmung. Mit steigender Vertrautheit mit dem Spiel stieg auch die Attraktivität (Bewertung im Rahmen des abschließenden Fragebogens auf einer 5-Punkte Likert-Skala: Session 1: 3,9 – Session 2: 4,5 – Session 3: 4,8). Aufbauen konnte das Szenario auf ein hohes Vorwissen, da über 70% der Teilnehmer SimCity vorher schon privat gespielt hatten. Teilweise detailliert definierte Ziele (wie die Fertigstellung der Erschließung der Gewerbegebiete) halfen in jedem Szenario ebenfalls, schnell Erfolgserlebnisse zu erzeugen und damit die Fokussierung auf das Spielgeschehen zu fördern. Eine ideale Gruppengröße wird sowohl durch die Befragung und Gruppendiskussion als auch durch die Beobachtung bei 3-4 Personen angesiedelt. Eine solche Größe erlaubt auf der einen Seite intensive Reflektion, auf der anderen Seite müssen sich auch alle Mitglieder einbringen.

4.2 Motivation und Lernen

In der Gruppendiskussion äußerten sich die Studierenden positiv über die verschiedenen Perspektiven, die durch die Simulation erläutert werden. Sie empfanden den Umgang mit den Problemstellungen als einfacher gegenüber traditionellen Lernmethoden, äußerten jedoch Zweifel, ob diese Methode zu einem vergleichbaren Ergebnis in einem Test führen würde. Aus didaktischer Sicht wurde die Abwechslung, die durch die „Spielzeit“ entstanden ist, von den Studierenden als motivationsförderlich eingestuft. Inwieweit die Beschäftigung mit SimCity auch tatsächlich als Spielen empfunden wurde, sollte mit der über den Fragebogen ausgegebenen In-Game-Variante des *Game Experience Questionnaires (GEQ)* ermittelt werden [IKP13]. Auf ein spielähnliches Empfinden deuten der hohe Wert für *Positive Affect* und die niedrigen Ergebnisse in den Kategorien *Negative Affect* und *Tension* hin. Der geringe Wert in der Kategorie *Competence* (d.h. die Spielenden fühlen sich nicht ausreichend vorbereitet, um die Probleme des Spiels zu lösen) sowie der nicht übermäßig hohe Wert für *Challenge* (d.h. diese Probleme scheinen aber auch nicht als große Herausforderung aufgefasst zu werden) geben aus unserer Sicht die Zweifel der Studierenden wieder, ob die mit Hilfe des Spiels vermittelten Inhalte die notwendige Komplexität für einen Transfer in die Realität besitzen. Dieses ist in der Lehr-Lernforschung nicht unbekannt: Lernen wird immer mit etwas „das anstrengen muss, damit es Erfolg hat“ in Verbindung gebracht. Jedoch steht es dem Ansatz des spielbasierten Lernens entgegen, bei dem die mögliche Anstrengung durch intrinsische Motivation zumindest teilweise kompensiert wird. Bestätigt wird die als nicht hoch empfundene Herausforderung durch einen Mittelwert von 3,2 (von 5) für 4 Items der Kategorie *Herausforderung* des FAM [RVB01]. Die im Vergleich hohe Bewertung von 4,1 in der Kategorie *Interesse* zeigt eine hohe Grundbereitschaft, sich mit dem Thema auseinanderzusetzen, was auch qualitativ in der Gruppendiskussion herausgearbeitet werden konnte. Bestandteil des Fragebogens war auch eine Gruppe von Aussagen, zu denen der Grad der Zustimmung (gleichfalls 5-Punkte Likert Skala) angegeben werden musste. Die Ergebnisse bestätigen teilweise die oben beschriebenen Feststellungen. Tendenziell wird das Spielen eher als Entspannung denn als Lernen empfunden ($\bar{x}=3,3$), eine Wirkung die mit dem Einsatz eines Spieles beabsichtigt ist. Es wird den Aussagen widersprochen, dass die Sachverhalte des Spiels nicht verstanden wurden (1,4) und dass die Aufgabe eher schwierig war (1,9). Lernförderliche „Aha-Erlebnisse“ wurden nur zu einem geringen Grad gesehen (2,7), da ein Großteil der Lerninhalte schon bekannt war (3,9). Dennoch wird nur von einer mittleren mentalen Anstrengung (3,0) ausgegangen. Auch werden teilweise neue Erkenntnisse (3,0) konstatiert. Bemerkenswert ist das hohe Interesse an einer Wirkungsbeobachtung (4,3) und deren Ergründung (3,4). Dieses kann als Alleinstellungsmerkmal des interaktiven Experimentierwerkzeugs *Spiel* interpretiert werden.

4.3 Herausforderung *Game-Aging*

Eine der wesentlichen Herausforderungen für den Einsatz eines Spiels über einen langen Zeitraum ist eine Erscheinung, die wir mit *Game-Aging* bezeichnen: ein Spiel altert über

die Zeit des Einsatzes. Diese äußert sich in unterschiedlichen Dimensionen. Als erstes ist *technische Alterung* zu nennen: Soft- und Hardware, mit denen die Spielsoftware zusammenarbeitet, ist nach einiger Zeit nicht mehr verfügbar. Eine Lösungsmöglichkeit ist der Einsatz von virtuellen Maschinen: durch den Einsatz von Virtualisierungssoftware – wie hier Oracle’s VirtualBox – kann der Zeitraum der Nutzbarkeit für die Spielsoftware bedeutend verlängert werden. Gleichfalls ist auch eine *kulturelle Alterung* zu beobachten: neue Spiele zeigen neue Spielelemente, Spielmechaniken oder auch Bedienmetaphern. Dadurch wird das eingesetzte Spiel klar als alt erkennbar. Dies kann Spielerlebnis und –akzeptanz senken. *Fachliche Alterung* liegt vor, wenn aktuelle fachliche Entwicklungen nicht mehr berücksichtigt sind.

5 Implikationen für Praxis und Forschung

Seit Jahren wird SimCity in Bildungskontexten eingesetzt. Der hier vorgestellte Anwendungsfall weist gegenüber vielen literaturbekannten Einsätzen mehrere Besonderheiten auf: (1) Er thematisiert ein regionales Planungsszenario (2) im Rahmen der universitären Hochschullehre über (3) eine lange Einsatzzeit von inzwischen fast 15 Jahren. (4) Methodisch hervorzuheben ist das offene, simultane und kompetitive Spiel in Gruppen.

Aus der Spielbeobachtung ergab sich ein relativ hohes Engagement der Kursteilnehmer. Dieses wurde gefördert durch die individuelle Betreuungsleistung mit hoher Detailkenntnis des Dozierenden. Der Einsatz des Spiels ist engmaschig strukturiert, was zu einer äußerst fokussierten Arbeitsweise führte. Gleichzeitig haben die Kursteilnehmer in ihren Spielhandlungen Freiheit und können das Szenario unter kurzfristigen Rückmeldungen wettbewerbsorientiert entwickeln. Das führte zu hohem Engagement. Die Spielsoftware wurde zwar als alt erkannt, blieb dadurch jedoch bedienbar. Eine fachliche Alterung der Szenarien wurde nicht festgestellt. Motivierende Wirkungen des regionalen Szenarios konnte nicht nachgewiesen werden, ist aber weiterhin aufgrund der Beobachtung des Seminars während den Spielszenarien zu vermuten. Zusätzlich waren die vorhandenen Detailkenntnisse bei der Entwicklung des Szenarios hilfreich. Gleichfalls wurde die Attraktivität in den Gruppendiskussionen bekundet.

Die Untersuchungen haben gezeigt, dass die Spieler zum einen engagiert und aufmerksam waren, die relevanten Sachverhalte jedoch nicht als überraschend, neu oder schwierig wahrgenommen haben. Dieses steht zunächst einmal in Widerspruch zu der Annahme, dass hinter der technischen Infrastruktur in SimCity ein komplexes Modell mit vielen Interdependenzen steht. Eine mögliche, zu validierende Erklärung ist, dass der Spielcharakter für die verzerrte Wahrnehmung der eigentlich schweren Aufgaben führt. Ebenfalls zu klären ist die Nachhaltigkeit der Lernerfahrungen. Zunächst entsteht durch den Wettbewerb Zeitdruck, der sich negativ auf Lernprozesse auswirken könnte. Dieser Zeitdruck - ausgelöst durch inhaltliche Faktoren wie komplexe Vernetzung aller Prozesse und Infrastrukturelemente, kontinuierlich weiterlaufende Systementwicklung, während noch Entscheidungsprozesse über sinnvolle Veränderungen der Rahmenbedingungen laufen - gehört jedoch auch zu den angestrebten Lernzielen. Des Weiteren ist es

untersuchenswert, ob auch entsprechende tiefe Lernerfahrungen ausgelöst werden. Darüber hinaus ist eine weitere Aufarbeitung des Themas *Game-Aging* notwendig. Eine systematische Betrachtung der Auswirkungen sollte genutzt werden, um zu einem konstruktiven Umgang zu gelangen.

Literaturverzeichnis

- [BPM10] Brünken, R.; Plass, J. L.; Moreno, R.: Current Issues and Open Questions in Cognitive Load Research. In (Plass, J. L.; Moreno, R.; Brünken, R. Hrsg.): *Cognitive Load Theory*. Cambridge University Press, Cambridge, S. 253–272, 2010.
- [Dö81] Dörner, D.: Über die Schwierigkeiten menschlichen Umgangs mit Komplexität. *Psychologische Rundschau XXXI*, S. 163–179, 1981.
- [Ga07] Gaber, J.: Simulating Planning. SimCity as a Pedagogical Tool. *Journal of Planning Education and Research* 27, S. 113–121, 2007.
- [HAB05] Haupt, T.; Arnold, U.; Bidlingmaier, W.: Studien- und hochschulübergreifender Einsatz einer engl.-spr. multimedialen Urban Infrastructure Development Simulation in der akademischen Aus- und Weiterbildung – MUrIDS: Tagungsband, 9. Workshop: Multimedia in Bildung und Wirtschaft – Einsatz und Nachhaltigkeit von eLearning, 22.-23.9.05 Ilmenau, Ilmenau, 2005.
- [IKP13] IJsselsteijn, W. A.; Kort, Y. A. W. de; Poels, K.: The Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games. Manuscript in Preparation, 2013.
- [Ma16] Mayer, R. E.: What Should Be the Role of Computer Games in Education? In *Policy Insights from the Behavioral and Brain Sciences* 3, S. 20–26, 2016.
- [RVB01] Rheinberg, F.; Vollmeyer, R.; Burns, B. D.: QCM. A questionnaire to assess current motivation in learning situations. *Diagnostica* 47, S. 57–66, 2001.
- [Sq03] Squire, K. R. et al.: Design Principles of Next-Generation Digital Gaming for Education. *Educational Technology* 43, S. 17–23, 2003.
- [Sw06] Sweller, J.: 1. How the human cognitive system deals with complexity. In (Elen, J.; Clark, R. E., Hrsg.): *Handling complexity in learning environments: Theory and research*. Elsevier, Amsterdam, S. 13–25, 2006.
- [Wo15] Woessner, M.: Teaching with SimCity. Using Sophisticated Gaming Simulations to Teach Concepts in Introductory American Government. *PS: Political Science & Politics* 48, S. 358–363, 2015.

Eine extravertierte und eine gewissenhafte Person in jeder Lerngruppe! Effekte der Verteilung von Persönlichkeitsmerkmalen auf Zufriedenheit und Lernergebnis

Henrik Bellhäuser¹, Johannes Konert², René Röpke³ und Christoph Rensing⁴

Abstract: Die Nutzung digitaler Werkzeuge in der Lehre erlaubt neben Unterstützung und Entlastung auch qualitative Verbesserungen. Eine algorithmische Unterstützung erlaubt die qualitative Verbesserung der Lerngruppenzusammensetzung mit dem Ziel, dass jeder Lernende den bestmöglichen Lernzuwachs erzielen kann und die Lerngruppen stabil und produktiv bleiben. Die Auswahl und Gewichtung relevanter Kriterien stellt dabei eine interdisziplinäre Herausforderung dar. Dieser Beitrag stellt den Stand der aktuellen Forschung zu algorithmischen Ansätzen, sowie zu relevanten Kriterien für Lerngruppenformation vor. Darauf aufbauende Schwerpunkte sind das Untersuchungsdesign, Evaluation und Ergebnisse zum Einfluss der Persönlichkeitsmerkmale Extraversion und Gewissenhaftigkeit auf die Produktivität, Zeitinvestment, Qualität der Ergebnisse und Zufriedenheit mit der Lerngruppe. Die Ergebnisse der durchgeführten Studie vom Herbst 2016 werden in Bezug gestellt zu den signifikanten Ergebnissen der Vorjahresstudie, welche Vorwissen und Motivation mit einbezog. Das Fazit des Beitrages zeigt, dass beide Persönlichkeitsmerkmale als heterogene Kriterien verteilt die besten Effekte hervorbringen.

Keywords: Lerngruppenformation, CSCL, Homogenität, Heterogenität, Extraversion, Gewissenhaftigkeit, Vorwissen, Motivation, MoodlePeers

1 Einleitung

Bei reiner Präsenzlehre und kleinen Kursen, können Lehrende anhand ihrer guten Kenntnis von Kompetenzständen und Interessen der Lernenden für bestimmte Lernziele optimierte Lerngruppen bilden und so optimale Bedingungen zum kollaborativen Kompetenzzuwachs schaffen. Die manuelle Gruppenzusammenstellung wird jedoch schnell zu komplex, wenn die Zahl der Lernenden steigt und die Kenntnis über deren Wissenstand und Kompetenzen begrenzt ist. In Blended Learning Szenarien oder reiner Online-Lehre ist Gruppenbildung ohne algorithmische Unterstützung praktisch nicht durchführbar. Gängige Lösung sind die randomisierte Zuweisung oder die

¹ Universität Mainz, Psychologie in den Bildungswissenschaften, Binger Str. 14-16, 55099 Mainz, bellhaeuser@uni-mainz.de

² Beuth Hochschule für Technik, FB VI Informatik und Medien, Luxemburger Str. 10, 13353 Berlin, johannes.konert@beuth-hochschule.de

³ RWTH Aachen University, Informatik 9 (Learning Technologies), Ahornstraße 55, 52074 Aachen, roepke@informatik.rwth-aachen.de

⁴ TU Darmstadt, FG Multimedia Kommunikation, Rundeturmstraße 10, 64283 Darmstadt, christoph.rensing@kom.tu-darmstadt.de

Gruppenbildung durch die Lernenden selbst. Beide Möglichkeiten haben wesentliche Nachteile [Mi12] und optimieren in keiner Weise die Zusammenstellung für das angestrebte Lernziel. Aus algorithmischer Sicht existieren inzwischen diverse Lösungsansätze, die das Problem der Lerngruppenformation adressieren. Einige Ansätze nutzen Machine Learning, um auf Basis von bisheriger Zusammenarbeit optimiert Lerngruppen zu bilden. Andere Verfahren setzen auf die Nutzung von Optimierungsverfahren auf Basis der Merkmalsvektoren der Lernenden und passenden Randbedingungen. Neben der Frage des passenden algorithmischen Ansatzes stellt sich auch die Frage, welche Kriterien für die Lerngruppenbildung relevant sind und auf welches Optimierungsziel sie Einfluss haben. Kriterien lassen sich dabei anhand ihrer Stabilität einordnen. Instabile Kriterien ergeben sich aus einer Relation zum aktuellen Lernszenario und den bearbeiteten Aufgaben. Beispielsweise der Anteil an Diskussionsbeteiligung oder die Bewertung eines Gruppenpartners nach der Aufgabebearbeitung. Diese Kriterien sind bei Nutzung digitaler System (bspw. Lernmanagementsysteme (LMS) wie Moodle⁵) leicht zu erfassen, doch ihre Relevanz für die Qualität der Gruppenarbeit ist nur eingeschränkt auf andere Szenarien oder Konstellationen übertragbar. Als stabile Kriterien gelten personenbezogene Merkmale, die sich auch in anderen Lernszenarien und Gruppenzusammensetzungen nicht stark ändern (z.B. Persönlichkeitsmerkmale). Erkenntnisse über deren Auswirkung auf die Qualität der Lerngruppe würden sich wiederholt in unterschiedlichen Szenarien untersuchen lassen und eine Verallgemeinerung und Übertragbarkeit der Relevanz für andere Kulturkreise und Lernszenarien erleichtern. Aufgrund der besseren Verallgemeinerbarkeit wird in der pädagogischen Psychologie insbesondere erforscht, welche stabilen Kriterien sich für die optimierte Lerngruppenformation nutzen lassen und wie hoch deren Einfluss auf die Qualität der Lerngruppen ist.

Während in der wissenschaftlichen Literatur von einer weitgehenden Stabilität der sogenannten Big Five-Persönlichkeitsmerkmale ausgegangen wird [MC94], existieren weiterhin nur wenige Studien darüber, welchen Einfluss diese auf die Lerngruppenqualität haben und ob einzelne oder mehrere der fünf Merkmale eher homogen (gleich) oder heterogen in einer Lerngruppe verteilt sein sollten. Nachdem eine dieser Arbeit vorausgegangene Studie der Autoren den signifikanten Einfluss der Kombination von Gewissenhaftigkeit und Extraversion, Motivation und Vorwissen auf die Qualität der Gruppenarbeit und die Zufriedenheit mit der Lerngruppe ergab [Rö16], ist das Ziel der hier vorgestellten Studie den Einfluss der beiden Persönlichkeitsmerkmale genauer zu erforschen, um das Verhältnis und die Wichtigkeit im Vergleich zu Vorwissen und Motivation zu bestimmen.

Darüber hinaus wird im Folgenden die hohe Qualität des Studiendesigns erörtert, um als Orientierung für weitere Studien zu gelten, die den Einfluss einer Intervention auf Qualitätsmerkmale, wie den Lernzuwachs, Zufriedenheit oder Ergebnisqualität zu untersuchen. Es folgt im nächsten Kapitel zunächst ein Überblick über die algorithmischen Herausforderungen, gefolgt von der Diskussion des Erkenntnisstandes

⁵ <http://www.moodle.org>, abgerufen am 03.03.2017

der pädagogischen Psychologie, bevor das Studiendesign beschrieben wird. Anschließend werden die Ergebnisse dargelegt, diskutiert und abschließend in Bezug zu den Ergebnissen der vorhergehenden Studie gesetzt.

2 Verwandte Arbeiten zur Lerngruppenformation

Für das Ziel der optimierten Lerngruppenformation ist es relevant, das Problem dieser Gruppenbildung von der Teambildung im organisationalen Kontext abzugrenzen. Einige der verwandten Arbeiten wenden Ansätze zur Lösung des Fertigungsinsel-Problems⁶ oder aus der Kapazitätsplanung an. Diesen Lösungen ist gemein, dass sie hinsichtlich des Ergebnisses optimieren, also der Ausgabe des Arbeitsprozesses unter möglichst optimaler Kombination der Ressourcen. Aus didaktischer Sicht geht es bei der Lerngruppenbildung aber gerade nicht nur um die Optimierung des Produktes, sondern um die Optimierung des Lernzuwachses jedes einzelnen Lernenden. Darüber hinaus spielt aus didaktischer Sicht die möglichst gleiche Qualität aller gebildeten Gruppen eine wesentliche Rolle, um jedem Lernenden die gleichen Lernzuwachs-Chancen zu bieten.

2.1 Computer-gestützte Lerngruppenformation (CSLGF⁷)

Die Berechnung einer optimalen Lösung übersteigt schon bei Kursgrößen unter 40 Personen die praktische Berechenbarkeit. Das Problem liegt in der Komplexitätsklasse von $\Omega(\alpha^M)$ mit $\alpha > 1$ und M der Anzahl an Teilnehmenden [Ko14, S. 90f]. Die Notwendigkeit zur Entwicklung guter algorithmischer Näherungslösungen wurde insbesondere für die Bereiche Tutorieller Systeme und web-basierter Trainings gefordert [WP01]. Die Vorteile sind inzwischen auch für Blended Learning, MOOCs und die Hochschullehre erkennbar. Vorhandene algorithmische Ansätze variieren in ihren Methoden (Randbedingungen oder numerische Optimierung), die Arten und Kombinationen der Kriterien (bspw. homogen oder gemischt) und der Häufigkeit der Gruppenbildung (einmalig, wiederholt) [vgl. SB14].

Die Randbedingung-basierten Ansätze formalisieren das Problem als Constraint Satisfaction Problem (CSP). Auf Basis von Regeln für gültige Lerngruppen und den Lernereigenschaften werden semantische Technologien, vorzugsweise Ontologien, verwendet, um gültige Lösungen zu finden. [ODM08] definieren eine erweiterte friend-of-a-friend Ontologie und unterstützen Restriktionen zu Gruppenrollen, Geschlecht und Gruppengröße bevor ein allgemeiner Logik-Löser (DLV) angewandt wird. [AOT12] modellieren die Kompetenzen der Lernenden und wenden ein Agenten-basiertes System an. Nachteilig bei agentenbasierten System ist die fehlende Berücksichtigung einer gleichförmigen Gruppenqualität neben der Optimierung auf die Ziele einzelner Agenten. Allen semantischen Ansätzen gemein ist die Notwendigkeit einer abgestimmten

⁶ In Englisch: Cell Formation Problem

⁷ In Englisch: Computer Supported Learning Group Formation

Ontologie, wodurch die Flexibilität in der Wahl der Kriterien, ebenso wie die Anwendung in anderen Lernszenarien eingeschränkt ist. Darüber hinaus liefern diese Ansätze keine Qualitätsmetrik, die einen Vergleich der Formationsqualität und Ähnlichkeit leicht zulässt.

Die numerischen Ansätze modellieren Merkmale der Studierenden und kontext-abhängige Merkmale als n-dimensionale Kriterienvektoren. Die Algorithmen von [Pa10] sowie [CP07] nutzen Clusteranalyse für die Gruppenformation, aber erlauben nicht die Berücksichtigung von homogenen und heterogenen Kriterien gleichzeitig. [GB06] nutzen Ameisenkolonie-Optimierung um in erster Linie heterogene Gruppen zu bilden. OmadoGenesis [Go07] erlaubt die Bildung homogener, heterogener und gemischter Gruppen, erlaubt jedoch nur diskrete Kriterienwerte und nutzt bei gemischten Kriterien (nur) einen genetischen Algorithmus. bei ihrem heuristischen, iterativen Verfahren. Nur zwei der in Literatur gefundenen Ansätze liefern originäre Algorithmen zur Erreichung nahezu optimaler Lösungen durch ziel-orientierte Vorsortierung der Lernenden und Auswahl nächster Gruppenmitglieder durch paarweise Vergleiche oder Vergleich mit Pivotelementen [GB06], [KBS14]. Da von diesen beiden GroupAL bereits im Vergleich zu verwandten Arbeiten besser Gruppenformationsergebnisse lieferte, wurde dieser als Grundlage für die im Folgenden dargestellte Evaluationsstudie verwendet.

2.2 MoodlePeers und GroupAL: Moodle-GUI und Optimierungsalgorithmus

GroupAL nutzt ein numerisches Verfahren auf Basis von n-dimensionalen Vektoren je Kriterium. Die gewichteten Kriterien können jeweils homogen oder heterogen optimiert werden. Dazu wird eine Abstandsmetrik verwendet, die paarweise die Passung (homogen) oder Ergänzung (heterogen) der Lernenden in den Kriteriendimensionen berechnet. Auf Basis des Abstandes wird die Eignung für die Gruppe bestimmt. Weitere Details können [Ko14] entnommen werden. Für die Nutzung in der Praxis wurde an der TU Darmstadt ein Plugin für Moodle entwickelt, welches die Kriterien für drei gängige Lernszenarien vorauswählt und Lehrenden so ein einfach zu bedienendes Werkzeug an die Hand gibt. Details sind [Rö16] zu entnehmen. Das Plugin ist als OpenSource Projekt⁸ unter GNU General Public License⁹ veröffentlicht und inzwischen in mehreren Versionen bei Moodle.org¹⁰ verfügbar.

2.3 Erkenntnisse der pädagogischen Psychologie

Kooperatives Lernen ist in einer Vielzahl empirischer Studien als äußerst wirkungsvolle Methode nachgewiesen worden [Ky13]. Die Vorteile gegenüber individuellem Lernen betreffen neben subjektiven Maßen wie Zufriedenheit auch objektive Leistungen. Weniger intensiv beforscht ist in diesem Zusammenhang die Frage, nach welchen

⁸ <https://github.com/moodlepeers>, abgerufen am 05.03.2017

⁹ <https://www.gnu.org/licenses/gpl-3.0.en.html>, abgerufen am 24.03.2017

¹⁰ https://moodle.org/plugins/mod_groupformation, abgerufen am 05.03.2017

Kriterien die Lerngruppenmitglieder ausgewählt werden sollten. In korrelativen Studien konnte gezeigt werden, dass demografische Eigenschaften wie Geschlecht, Alter oder Bildungsniveau verhältnismäßig gering mit dem Erfolg von Gruppen zusammenhängen [Ha02]. Demgegenüber fanden sich stärkere Zusammenhänge zwischen psychologischen Merkmalen wie Persönlichkeits-eigenschaften, Einstellungen und Gruppenleistungen.

Von den fünf Dimensionen der weitverbreiteten Big Five-Persönlichkeitseigenschaften [RJ05] sind es vor allem Extraversion und Gewissenhaftigkeit, die als relevant für die Gruppenformation gelten [Hu07]. Für Extraversion – eine Eigenschaft, die neben Geselligkeit auch Führungsverhalten als Facette enthält – gilt dabei eine heterogene Gruppenzusammensetzung als förderlich, um Konflikte zwischen mehreren Personen mit Führungsanspruch zu vermeiden. Für Gewissenhaftigkeit wird eine homogene Gruppenzusammensetzung als förderlich betrachtet, da größere Differenzen zwischen den Gruppenmitgliedern hinderlich für die Formulierung gemeinsamer Gruppenziele wären; Mitglieder mit größerer Gewissenhaftigkeit würden vermutlich höhere Ansprüche an die Gruppenleistung stellen, als dies bei Mitgliedern mit niedriger Gewissenhaftigkeit der Fall wäre. Für eine einzelne Gruppe wäre zudem wünschenswert, dass die Gewissenhaftigkeit in der Gruppe nicht nur homogen, sondern auch im Durchschnitt hoch ausgeprägt ist, da Gewissenhaftigkeit für Einzelpersonen als guter Prädiktor für Leistungen gilt [RAB12]. Allerdings kann bei einer gegebenen Population, die vollständig in Gruppen eingeteilt werden muss, der Durchschnittswert einer Gruppe nur auf Kosten der anderen Gruppen erhöht werden. Aus Gründen der Fairness kann daher nicht das Gruppenniveau optimiert werden.

Abseits der Big Five-Persönlichkeitsdimensionen wurden diverse andere psychologische Merkmale als relevant für Gruppenformationen beobachtet. [Be07] benennt Teamorientierung als Kriterium, das homogen verteilt sein sollte, damit sich die Gruppenmitglieder einig über den Kollaborationsgrad sind. Weitere diskutierte Faktoren sind Motivation und Zielorientierung [NVV11], sowie allgemeine kognitive Fähigkeiten und Vorwissen [Ho05].

3 Forschungsfragen und Fokus der Studie

Aufgrund guter psychometrischer Eigenschaften und der gut zugänglichen Messinstrumente liegen mehr Studien zum Einfluss von Extraversion und Gewissenhaftigkeit auf die Gruppenformation vor als zu den anderen Einflussfaktoren. Als Limitation der empirischen Befundlage ist zu nennen, dass die Erkenntnisse aus rein korrelativen Studien stammen – typischerweise wurden dabei Gruppen untersucht, die entweder per Zufall gebildet wurden oder durch die Studierenden selbst. Qualitativ höherwertige Studiendesigns, insbesondere randomisierte Experimentalstudien, die den Rückschluss auf kausale Effekte erlauben, fehlen bisher weitgehend. Eine Ausnahme stellt eine von den Autoren dieses Beitrages durchgeführte Pilotstudie dar, in der gezeigt werden konnte, dass die algorithmische Berücksichtigung einer Kombination von mehreren

relevanten Persönlichkeitseigenschaften statistisch signifikant überlegen war in Bezug auf die Zufriedenheit und Gruppenleistung im Vergleich zu einer reinen Zufallsgruppierung [Rö16]. Allerdings ließ sich nachträglich nicht beurteilen, welchen relativen Anteil die einzelnen Kriterien am positiven Gesamteffekt hatten und ob gegebenenfalls Interaktionen zwischen den Kriterien auftraten.

In der vorliegenden Studie soll der Einfluss von Extraversion und Gewissenhaftigkeit systematisch untersucht werden. Dabei sollen die Effekte beider Kriterien sowohl separat voneinander als auch in ihrer Kombination analysierbar sein. Da die gleichzeitige Berücksichtigung von zwei Kriterien eine größere Restriktion für den Gruppenformationsalgorithmus darstellt als die Berücksichtigung nur eines Kriteriums, wird erwartet, dass bei der Kombination ein negativer Interaktionseffekt eintritt. Um besonders robuste Effekte zu erzielen, sollen nicht nur die für Extraversion und Gewissenhaftigkeit jeweils wünschenswerten heterogenen bzw. homogenen Formationsbedingungen realisiert werden, sondern auch die jeweils theoretisch unerwünschten Bedingungen (d.h. homogene Extraversion und heterogene Gewissenhaftigkeit). Zum Vergleich sollen auch Formationsbedingungen herangezogen werden, bei denen jeweils nur ein Kriterium verwendet und das andere ignoriert (d.h. zufällig behandelt) wird sowie eine reine Zufallsbedingung. Aufgrund der umfangreichen Analyseoptionen, die sich aus dem beschriebenen Forschungsdesign ergeben, wird im Folgenden nur eine Teilmenge der möglichen Analysen dargestellt. Der Fokus liegt dabei auf der gleichzeitigen Verwendung beider Persönlichkeitseigenschaften; Varianten, in denen jeweils nur eine davon für die Gruppenformation verwendet werden, werden hier ausgespart.

3.1 Forschungsfragen

Auf Basis der empirischen Befundlage werden folgende Hypothesen untersucht:

- H1) Gruppen, bei denen algorithmisch eine heterogene Verteilung der Extraversion hergestellt wird, berichten eine größere Zufriedenheit mit der Gruppenzusammenstellung und der Zusammenarbeit, investieren mehr Zeit auf die Gruppenarbeit und erzielen bessere Ergebnisse als Gruppen, bei denen eine homogene Verteilung der Extraversion hergestellt wird.
- H2) Gruppen, bei denen algorithmisch eine homogene Verteilung der Gewissenhaftigkeit hergestellt wird, berichten eine größere Zufriedenheit mit der Gruppenzusammenstellung und der Zusammenarbeit, investieren mehr Zeit auf die Gruppenarbeit und erzielen bessere Ergebnisse als Gruppen, bei denen eine heterogene Verteilung der Gewissenhaftigkeit hergestellt wird.

3.2 Studiendesign und Durchführung

Zur Untersuchung der Forschungsfragen wurde ein zweifaktorielles Studiendesign verwendet. Sowohl der Faktor *Extraversion* als auch der Faktor *Gewissenhaftigkeit*

waren dabei dreifach gestuft mit den Ausprägungen *homogen*, *zufällig* und *heterogen*. Die daraus resultierenden neun Versuchsbedingungen werden in Tab. 1 dargestellt. Der Vorteil dieses vollständig gekreuzten Versuchsdesigns besteht darin, dass der jeweilige Einfluss von Extraversion (Haupteffekt 1) und Gewissenhaftigkeit (Haupteffekt 2) unabhängig voneinander untersuchbar ist. Zusätzlich kann so ein eventueller Interaktionseffekt der beiden Faktoren miteinander getestet werden, der über die bloße Addition der beiden Haupteffekte hinausgeht. Im Folgenden werden nur die Bedingungen 1, 3, 7 und 9 betrachtet und miteinander verglichen.

Tab. 1: Neun Versuchsbedingungen als Resultat des zweifaktoriellen Studiendesigns

		Gewissenhaftigkeit		
		heterogen	zufällig	homogen
Extraversion	heterogen	1	2	3
	zufällig	4	5	6
	homogen	7	8	9

Die Zuordnung zu den einzelnen Bedingungen erfolgte auf Zufallsbasis. Die Teilnehmenden konnten nicht erkennen, zu welcher Bedingung sie zugeordnet wurden und wussten nicht, welches das Ziel der Studie ist. Damit wurden Selektionseffekte oder erwartungskonformes Verhalten ausgeschlossen. Mit diesen Merkmalen entspricht das Experimentalstudiendesign der Qualitätsstufe 1+ nach [HM01]. Um die Überlagerung mit Effekten persönlicher Sympathie und eventuell vorher bestehenden Freundschaften zu minimieren, wurde als Szenario zur Durchführung ein Mathematik-Vorkurs der Universität für Erstsemester verschiedener Studiengänge genutzt. Die Teilnehmenden nahmen ausschließlich online über Moodle über vier Wochen an dem freiwilligen und unbenoteten Kurs teil. Aufgrund der Freiwilligkeit und der fehlenden Relevanz für die Benotung im Studium war aus ethischen Gründen die Zuordnung zu den Experimentalbedingungen möglich. Andernfalls wäre eine Aufklärung der Teilnehmenden und eine Auswahlmöglichkeit von Nöten, was die Qualität des Experimentes schwächt. Aufgrund der bislang noch schwachen empirischen Fundierung erscheint zudem eine Einteilung in zum Teil ungünstige Gruppenkonstellationen als zumutbar. Erst wenn der Nachweis von leistungsförderlichen bzw. -hinderlichen Gruppenformationen in qualitativ hochwertigen Studien erbracht wurde, dürfen Versuchspersonen nicht mehr bewusst in für sie schädliche Studienbedingungen gebracht werden. Dennoch hatten alle Teilnehmenden im vorgestellten Szenario die Möglichkeit ihre Zustimmung zur Teilnahme an der Studie zu verneinen (opt-in mit textueller Datenschutzaufklärung). Sie wurden dann in zufällige Gruppen eingeteilt, die nicht Bestandteil der Kontrollgruppen waren. Als Gruppengröße wurden vier Mitglieder pro Gruppe angestrebt; in Ausnahmefällen wurden auch drei Mitglieder erlaubt. Diese Entscheidung fiel nach Abwägung von zwei Aspekten: Einerseits sollte eine Gruppe den Ausfall eines Mitglieds kompensieren und arbeitsfähig bleiben können; andererseits sollte die Gesamtanzahl von gebildeten Gruppen maximiert werden. Im Rahmen des Vorkurses erhielten die Teilnehmenden insgesamt drei Gruppenaufgaben, die jeweils im

Laufe einer Woche zu bearbeiten waren. Die Aufgaben wurden von Mathematikdidaktikern als offene Modellierungsaufgaben ausgearbeitet, sodass die Problemlösekompetenz geschärft und Kollaboration gefördert wurde.

Als Kontrollvariablen wurden über das MoodlePeers-Plugin neben dem Fragebogen zu den Big Five [RJ05] auch die Dimensionen Selbstregulationskompetenz [Be16], Motivation [RVB01] und Selbstwirksamkeit [SJ99] und demografische Daten erhoben. Nach Abschluss des Vorkurses beantworteten die Teilnehmenden einen Evaluationsbogen. Dieser enthielt Fragen zur Zufriedenheit mit der Gruppenzusammenstellung und zur selbsteingeschätzten Produktivität der Gruppe (jeweils Likert-skaliert von 1 bis 6) sowie die Frage, wie viel Zeit wöchentlich auf die Gruppenarbeit verwendet wurde. Als objektives Maß für die Qualität der Gruppenarbeit wurde ausgezählt, wie viele der drei Gruppenaufgaben erfolgreich bearbeitet wurden.

Die Studie wurde mit 660 Teilnehmenden des Vorkurses im Zeitraum von 12.09.2016 bis 07.10.2016 durchgeführt von denen 461 Studierende der Teilnahme zustimmten und vollständige Daten für die Gruppenformation abgaben. Die Aufteilung in die Experimentalbedingungen ist in Tab. 2 mit angegeben.

Tab. 2: Aufteilung der Probanden auf die Experimentalbedingungen

		Gewissenhaftigkeit		
		heterogen	zufällig	homogen
Extraversion	heterogen	n Gruppen = 14 n Individuen = 55	n Gruppen = 13 n Individuen = 52	n Gruppen = 14 n Individuen = 55
	zufällig	n Gruppen = 12 n Individuen = 48	n Gruppen = 14 n Individuen = 52	n Gruppen = 12 n Individuen = 48
	homogen	n Gruppen = 13 n Individuen = 52	n Gruppen = 12 n Individuen = 48	n Gruppen = 13 n Individuen = 51

4 Ergebnisse

Wie bei freiwilligen Lehrveranstaltungen üblich [Rö16], verzeichnete der Vorkurs einen erheblichen Dropout und nicht alle der verbliebenen Teilnehmenden beteiligten sich am Evaluationsfragebogen für die vorliegende Studie. Von den $n=159$ erhaltenen Bögen entfielen auf die Bedingungen 1, 3, 7 und 9 jeweils $n=20$, $n=15$, $n=16$ bzw. $n=11$. Für die erfolgreich bearbeiteten Hausaufgaben konnten hingegen fast alle Personen berücksichtigt werden (Bedingungen 1, 3, 7 und 9 jeweils $n=52$, $n=44$, $n=48$ bzw. $n=48$). Tab. 3 fasst die deskriptiven Ergebnisse für die abhängigen Variablen zusammen. Es ist ersichtlich, dass jeweils in Bedingung 1 (Extraversion und Gewissenhaftigkeit beide heterogen verteilt) die besten Ergebnisse zu verzeichnen sind – die Teilnehmenden berichten hier die größte Zufriedenheit mit der Zusammenstellung ihrer Gruppe,

schätzen die Produktivität ihrer Gruppe am höchsten ein, berichten das größte wöchentliche Zeitinvestment für die Gruppenaufgaben und bearbeiten auch objektiv den größten Anteil der Gruppenaufgaben.

Tab. 3: Deskriptive Ergebnisse

	Bedingung 1	Bedingung 3	Bedingung 7	Bedingung 9
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>
Zufriedenheit	4,11 (1,79)	2,46 (1,13)	3,00 (1,75)	3,40 (1,51)
Produktivität	4,00 (1,89)	1,62 (0,77)	2,31 (1,99)	2,30 (1,57)
Zeitinvestment	3,00 (3,43)	0,53 (0,92)	1,00 (1,36)	0,82 (1,33)
Hausaufgaben	0,92 (1,10)	0,45 (0,79)	0,39 (0,83)	0,35 (0,81)

Inferenzstatistische Analysen (2x2 ANOVA) ergaben dabei das folgende Bild:

Für die **Zufriedenheit mit der Gruppenzusammenstellung** ergab sich ein signifikanter Interaktionseffekt ($F(1, 54)=5,54$; $p=0,02$; $\eta^2=0,09$), während die Haupteffekte für Extraversion und Gewissenhaftigkeit nicht signifikant wurden. Demnach ist der Effekt der beiden unabhängigen Variablen abhängig von der Ausprägung der jeweils anderen Variable. So spielt es bei homogener Extraversion in einer Gruppe keine Rolle, ob die Gewissenhaftigkeit homogen oder heterogen verteilt ist; bei heterogener Extraversion ist hingegen heterogene Gewissenhaftigkeit deutlich vorteilhaft.

Für die **selbsteingeschätzte Produktivität** der Gruppe fanden wir einen signifikanten Haupteffekt für Gewissenhaftigkeit ($F(1, 54)=7,15$; $p=0,01$; $\eta^2=0,12$), sowie einen signifikanten Interaktionseffekt ($F(1, 54)=7,00$; $p=0,01$; $\eta^2=0,12$). Heterogene Gewissenhaftigkeit wirkt sich demnach grundsätzlich positiv aus und kann durch zusätzliche heterogene Extraversion noch verstärkt werden.

Für das **selbstberichtete Zeitinvestment** konnten ein signifikanter Haupteffekt für Gewissenhaftigkeit ($F(1, 54)=5,31$; $p=0,03$; $\eta^2=0,08$) sowie ein signifikanter Interaktionseffekt gezeigt werden ($F(1, 54)=3,96$; $p=0,05$; $\eta^2=0,06$). Auch hier ist heterogene Gewissenhaftigkeit grundsätzlich von Vorteil und wird durch zusätzliche heterogene Extraversion noch verstärkt.

Für die **erfolgreich bearbeiteten Gruppenaufgaben** zeigte sich ein signifikanter Haupteffekt für Extraversion ($F(1, 188)=5,40$; $p=0,02$; $\eta^2=0,03$) sowie ein signifikanter Haupteffekt für Gewissenhaftigkeit ($F(1, 188)=4,14$; $p=0,04$; $\eta^2=0,02$). Demnach ist heterogene Extraversion grundsätzlich vorteilhaft, ebenso wie heterogene Gewissenhaftigkeit. Die Kombination von beidem hingegen scheint keinen besonderen Effekt zu haben, der über die Addition beider Haupteffekte hinausgeht.

5 Diskussion und Fazit

Die Ergebnisse der Studie stellen teilweise eine Bestätigung der Hypothesen dar, stehen aber zu weiten Teilen im Kontrast zu ihnen. Für Hypothese 1 fand sich teilweise Bestätigung bei der erfolgreichen Bearbeitung der Gruppenaufgaben. Wie vermutet zeigte sich hier eine Überlegenheit von Gruppen mit heterogener Verteilung der Extraversion. In den anderen drei abhängigen Variablen hingegen konnten keine signifikanten Haupteffekte für Extraversion festgestellt werden. Dabei sollte aber nicht außer Acht gelassen werden, dass es sich dabei um subjektive Selbstberichtsdaten handelt, die zudem unter dem großen Dropout in der Stichprobe litten. Die signifikanten Interaktionseffekte den drei Selbstberichtmaßen können mit Einschränkungen ebenfalls als Bestätigung der Hypothese 1 betrachtet werden: Heterogene Extraversion wirkt sich hier positiv aus, sofern gleichzeitig Heterogenität in Bezug auf die Gewissenhaftigkeit in der Gruppe hergestellt wird.

Hypothese 2 muss auf Basis der gefundenen Ergebnisse hingegen verworfen werden. Nicht Homogenität, sondern Heterogenität in der Gewissenhaftigkeit scheint sich im Durchschnitt positiv auf das Lernen in Lerngruppen auszuwirken. Entsprechende signifikante Haupteffekte wurden für drei von vier abhängigen Variablen gefunden (Einschätzung der Produktivität der Gruppe, selbstberichtetes Zeitinvestment und objektive Gruppenarbeitsergebnisse). Anhand der berichteten Interaktionseffekte kann zudem gefolgert werden, dass sich der positive Effekt heterogener Gewissenhaftigkeit durch gleichzeitige Heterogenität in Extraversion verstärkt.

Die fehlende Übereinstimmung der Ergebnisse dieser Studie mit den Befunden aus der Literatur kann auf unterschiedliche Ursachen zurückzuführen sein, unter anderem auf Limitationen der vorliegenden Studie. Insbesondere die nach Dropout geringe Stichprobengröße stellt eine Bedrohung der externen Validität dar. Allerdings liegen zumindest für die Bearbeitung der Gruppenaufgaben vollständige Daten vor.

Untersucht wurde die Gruppenarbeit in einem außergewöhnlichen Setting mit ausschließlich virtueller Teamarbeit, die zudem nur in relativ niedriger Intensität durchgeführt wurde, bei insgesamt eher geringer extrinsischer Motivation der Teilnehmenden. Möglicherweise ist in einem solchen Szenario weniger die Homogenität oder Heterogenität der Gruppenmitglieder entscheidend, da sie sich zu wenig intensiv kennenlernen, um überhaupt Ähnlichkeiten oder Unterschiedlichkeiten beobachten zu können. Vielmehr könnte entscheidend sein, mindestens eine Einzelperson pro Gruppe zu haben, die die Gruppenarbeit initiiert und vorantreibt. Persönlichkeitseigenschaften, die ein solches Verhalten begünstigen, könnten eine hohe Gewissenhaftigkeit kombiniert mit hoher Extraversion darstellen. Diese Hypothese ist mit den hier gefundenen Ergebnissen gut vereinbar: Gruppen, die per Algorithmus sowohl heterogene Extraversion als auch heterogene Gewissenhaftigkeit zugeteilt bekommen, haben eine erhöhte Chance, solche Einzelpersonen zu enthalten.

Mangelnde Übereinstimmung mit älteren Befunden können aber unter Umständen auch

auf Schwächen jener zurückgeführt werden. So wurden Kompositionseffekte bei der Gruppenarbeit bislang ausschließlich korrelativ untersucht. Hier weist das von uns gewählte experimentelle Design einen klaren Vorteil auf und ist daher vorzuziehen. Im Vergleich mit den Ergebnissen unserer eigenen Pilotstudie [Rö16] kann geschlussfolgert werden, dass die positiven Ergebnisse der dort berichteten Kombination vieler Kriterien teilweise auf die dort hergestellte heterogene Verteilung von Extraversion zurückgeführt werden könnte. Allerdings wurde dort auch eine homogene Verteilung der Gewissenhaftigkeit verwendet, was auf Basis der neuen Befunde eher nicht zum Erfolg beigetragen haben dürfte. Dies legt nahe, dass die übrigen Kriterien (heterogenes Vorwissen, homogene Verträglichkeit, heterogener Neurotizismus, heterogene Offenheit, homogene Motivation und homogene Teamorientierung) einen größeren Anteil am Effekt hatten.

Literaturverzeichnis

- [AOT12] Abnar, S.; Orooji, F.; Taghiyareh, F.: An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. 2012 IEEE Int Conf Technol Enhanc Educ., S. 1-6, 2012.
- [Be07] Bell, S.T.: Deep-level composition variables as predictors of team performance: a meta-analysis. *J Appl Psychol* 92, S. 595-615, 2007.
- [Be16] Bellhäuser, H.; Lösch, T.; Winter, C.; Schmitz, B.: Applying a web-based training to foster self-regulated learning - Effects of an intervention for large numbers of participants. *Internet High Educ* 31, S. 87-100, 2016.
- [CP07] Christodoulopoulos, C.E.; Papanikolaou, K. A.: A Group Formation Tool in an E-Learning Context. 19th IEEE Int Conf Tools with Artif Intell 2007, S. 117-123, 2007.
- [GB06] Graf, S.; Bekele, R.: Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 4053 LNCS, S. 217-226, 2006.
- [Go07] Gogoulou, A. et al.: Forming Homogeneous, Heterogeneous and Mixed Groups of Learners. In (Brusilovsky, P., Hrsg): *Proc. Work. Pers. E-Learning Environ. Individ. Gr. Level*, 11th Int. Conf. User Model, S. 33-40, 2007.
- [HM01] Harbour, R.; Miller, J.: A new system for grading recommendations in evidence based guidelines. *Br Med J* 323, S. 334-6, 2001.
- [Ho05] Horwitz, S.K.: The Compositional Impact of Team Diversity on Performance: Theoretical Considerations. *Hum Resour Dev Rev* 4, S. 219-245, 2005.
- [Ha02] Harrison, D.A.; Price, K.H.; Gavin, J.H.; Florey, A.T.: Time, Teams, and Task Performance: Changing Effects of Surface- and Deep-Level Diversity on Group Functioning. *Acad Manag J* 45, S. 1029-1045, 2002.
- [Hu07] Humphrey, S.E.; Hollenbeck, J.R.; Meyer, C.J.; Ilgen, D.R.: Trait configurations in self-managed teams: a conceptual examination of the use of seeding for maximizing

- and minimizing trait variance in teams. *J Appl Psychol* 92, S. 885-892, 2007.
- [KBS14] Konert, J.; Burlak, D.; Steinmetz, R.: The Group Formation Problem: An Algorithmic Approach to Learning Group Formation. In (Rensing, C. et al, Hrsg): Proc. 9th Eur. Conf. Technol. Enhanc. Learn., Springer Berlin u.a., S. 221-234, 2014.
- [Ko14] Konert, J.: Interactive Multimedia Learning: Using Social Media for Peer Education in Single-Player Educational Games. Springer, Heidelberg, Germany, 2014.
- [Ky13] Kyndt, E. et al.: A meta-analysis of the effects of face-to-face cooperative learning. Do recent studies falsify or verify earlier findings? *Educ Res Rev* 10, S. 133-149, 2013.
- [MC94] McCrae, R.R.; Costa, P.T.: The Stability of Personality: Observations and Evaluations. *Curr Dir Psychol Sci* 3, S. 173-175, 1994.
- [Mi12] Mitchell, S.N.; et al.: Friendship and Choosing Groupmates: Preferences for Teacher-selected vs. Student-selected Groupings in High School Science Classes. *J Instr Psychol* 31, S. 1-6, 2012.
- [NVV11] Nederveen, Pieterse A.; van Knippenberg, D.; van Ginkel, W.P.: Diversity in goal orientation, team reflexivity, and team performance. *Organ Behav Hum Decis Process* 114, S. 153-164, 2011.
- [ODM08] Ounnas, A.; Davis, H.; Millard, D.: A Framework for Semantic Group Formation. Eighth IEEE Int Conf Adv Learn Technol, S. 34-38, 2008.
- [Pa10] Paredes, P.: A Method for Supporting Heterogeneous-Group Formation through Heuristics and Visualization 16, S. 2882-2901, 2010.
- [RAB12] Richardson, M.; Abraham, C.; Bond, R.: Psychological correlates of university students' academic performance. *Psychol Bull* 138, S. 353-87, 2012.
- [RVB01] Rheinberg, F.; Vollmeyer, R.; Burns, B.D.: FAM: Ein Fragebogen zur Erfassung aktueller Motivation QCM: A questionnaire to assess current motivation in learning situations. *Diagnostica* 2, S. 57-66, 2001.
- [RJ05] Rammstedt, B.; John, O.P.: Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit. *Diagnostica* 51, S. 195-206, 2005.
- [Rö16] Röpke, R.; et al.: MoodlePeers: Automatisierte Lerngruppenbildung auf Grundlage psychologischer Merkmalsausprägungen in E-Learning-Systemen. In (Lucke, U.; Schwill, A.; Zender, R., Hrsg.): Proc. der 14. E-Learning Fachtagung Inform. der GI, Köllen, S. 233-244, 2016.
- [SB14] Srba, I.; Bielikova, M.: Dynamic Group Formation as an Approach to Collaborative Learning Support. *IEEE Trans Learn Technol* 8, S. 173-186, 2014.
- [SJ99] Schwarzer, R.; Jerusalem, M.: Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen, 1999.
- [WP01] Wessner, M., Pfister, H.-R.: Group formation in computer-supported collaborative learning. In: Proc. 2001 Int. ACM Sigr. Conf. Support. Gr. Work - Gr. '01, S. 24-31, 2001.

Recomposing Small Learning Groups at Scale—A Data-driven Approach and a Simulation Experiment

Zhilin Zheng¹, Niels Pinkwart²

Abstract: Group re-composition has thus far been rarely studied. The recent emergence of large scale online learning contexts (e.g. MOOCs) might bring about an opportunity for its application due to the reported high drop-out rate. In this paper, we propose a novel data-driven approach to address the problem of group re-composition. Through a simulation experiment, we saw its capability in decreasing the drop-out rate in groups and bringing more cohesive groups when compared against a random grouping strategy.

Keywords: Group Formation; Group Re-composition; Group Dynamics; MOOC; Learning Analytics.

1 Introduction

Back in the 1950s, teachers and researchers began to experiment with within-class grouping, between-class grouping and even cross-grade grouping [Ga86, SI87, Ku92]. They often composed small learning groups according to students' achievement, attainments and aptitudes. More recently, other grouping features such as learning styles, demographic characteristics or behavioural attributes have been tested as well. Although much research attempts to find the generalizable theories that can make efficient groups in many cases, a consensus has not yet been reached on which type of groups are more likely to yield great learning outcome. In recent years, large-scale learning contexts, typically Massive Open Online Courses (MOOCs), bring about opportunities to study this problem in a new direction. Data-driven methods have begun to be employed [We16]. Basically, the data-driven methods create a data model that can predict grouping outcomes based on analysis of students' relevant data. This data model can then be used to make groups that could be more potentially successful. These methods obviously do not reapply the generalised grouping theories anymore. But the higher requirement of data size challenges their practical implementation. Obviously, too few data can hardly train a relatively unbiased data model. Nevertheless, massive enrolments in recent MOOCs, in principle, secure the quantity of data. Another typical feature in recent MOOCs, the reported high dropout rates [Jo14], could however make many small learning groups dysfunctional, because very scarce human resource would be left behind in those groups. Especially when we set up group learning in a multi-task context, recomposing those dysfunctional groups for the next group tasks would be necessary.

¹ Humboldt University of Berlin, Department of Computer Science, Berlin, zhilin.zheng@hu-berlin.de

² Humboldt University of Berlin, Department of Computer Science, Berlin, niels.pinkwart@hu-berlin.de

Thus, in the paper, we propose a data-driven approach to address the group re-composition problem. The remainder of this paper is organized as follows. After a review of recent studies on relevant topics, we propose our approach and report on a test using a simulation platform. Next, we interpret the experimental results and finally conclude our findings.

2 State of the art

As digitalized learning or e-learning technologies advanced, group composition has been studied on the basis of computer-supported methods. Among such, Moreno et al. [MOV12] suggested a genetic algorithm. Graf et al. [GB06] proposed an ant colony optimization method. Hsu et al. [Hs14] recommended an artificial bee algorithm and Zheng et al. employed a discrete particle swarm optimization approach [ZP14]. The grouping attributes employed vary and include Belbin roles [YA12], learning performance [GB06] and background knowledge [Hs14]. With regard to the grouping results, most of the approaches produce either homogeneous groups or heterogeneous groups or a mixture of both in some cases [KBS14]. All approaches rely on static data collected before a group task starts and none of those methods accounts for data alterations caused by group dynamics. For example, individual's roles or learning performance could change over time.

Group re-composition sometimes is unavoidable during group operation for several reasons. First, teachers may find dysfunction in groups. Second, students themselves want to dissolve their groups [SA10, p.55]. In classrooms, manual redistribution of those students would probably not be a bad choice. However, in large-scale online learning contexts, it would not be affordable anymore. To the best of our knowledge, only one recent publication attempts to address this group re-composition problem [SB14]. This paper proposes a dynamic group formation method to improve Computer Supported Collaborative Learning (CSCL) groups. The authors dynamically retrieved group interaction data and iteratively made use of it to compose groups for each new task. This work pioneers the use of group dynamics to compose groups. Yet, it still leaves much room for improvement. First, the authors merely observed the dynamic change of each student and reapplied a grouping method (i.e. Group Technology) to recompose learning groups over and over again, that is, the subtle association between group dynamics and group success has not been taken into account. Second, the method cannot offer us the concrete rules to compose either successful or weak groups. Instead, they repeatedly used the same grouping criteria (which may be questionable for any specific group). Third, it is necessary to scale up the method for recent larger-scale online learning platforms.

3 Methods

3.1 Group re-composition approach

An overview of our proposed algorithmic re-grouping method is depicted in Fig. 1. First of all, we can start with an initial group formation. Random groups can be employed here unless useful participant data is available. The goal is to improve this initial group formation in the following task(s).

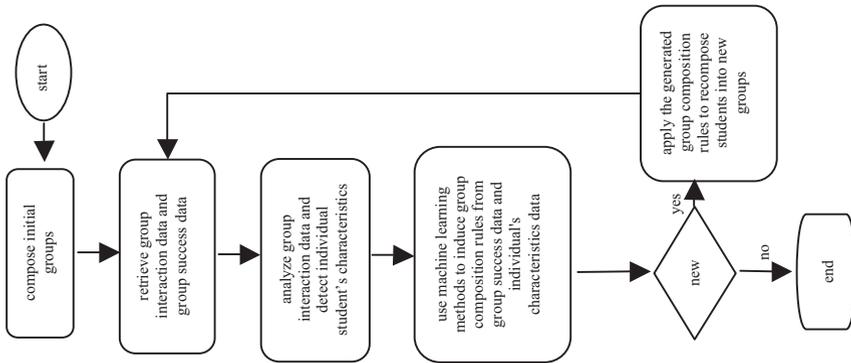


Fig. 1: Schema of the proposed group re-composition approach.

Next, we retrieve group data from the first group task. This data mainly comes from two sources: group interactions and group success (e.g. learning performance). Based on group interaction data, we can further detect individual student's roles in his/her group using Social Network Analysis (SNA). Merging with group success data, we can then apply machine learning methods to induce group composition rules that indicate which group roles (combined together) make successful groups or weak groups. Those generated composition rules are employed to suggest new groups for the next task if necessary. Through this iterated process, we can learn group composition knowledge from the data and apply it to recompose groups task by task.

The proposed approach has at least five advantages. First, it does not rely on potentially overly general grouping theories (that frequently do not apply to the situation of concrete groups). It is totally data-driven and directly reflects the truth encoded in the data. Second, it also works with students of whom no initial information is known. Third, it accounts for group dynamics. For example, if somebody's role changes over time, the data would accordingly reflect such. Fourth, it makes a dynamic connection between grouping attributes and group outcome. It could easily be extended to grouping attributes other than roles. Fifth, this approach is able to generate concrete group composition rules that can be applied to suggest group assignment in some other similar learning contexts.

3.2 Simulation system

Due to the nature of a data-driven approach, data source certainly plays an important role in evaluating the proposed approach. As shown in Fig. 1, group interaction data and group success data are two main data sources. Empirical methods such as field studies would be a straightforward way to collect such data. However, a suitable MOOC test course has not yet been found, perhaps, due to the requirement of designing multiple group tasks in a course. Instead, computer simulation was chosen to validate the proposed approach. Thanks to Nygren's studies [Ny10, Ny11, Ny12], modelling of group discussion is computationally possible. We thus chose it to generate the group interaction data in this paper. Via simulation, the group success data, such as learning performance, is not realistic to be collected either. We thus had to choose group cohesion that can be calculated from each simulating group as an alternative.

Our simulation system as a whole functionally consists of group composer, group interaction simulator, roles detector, machine learning and regrouping components. Fig. 2 gives an overview on how these five main components work together.

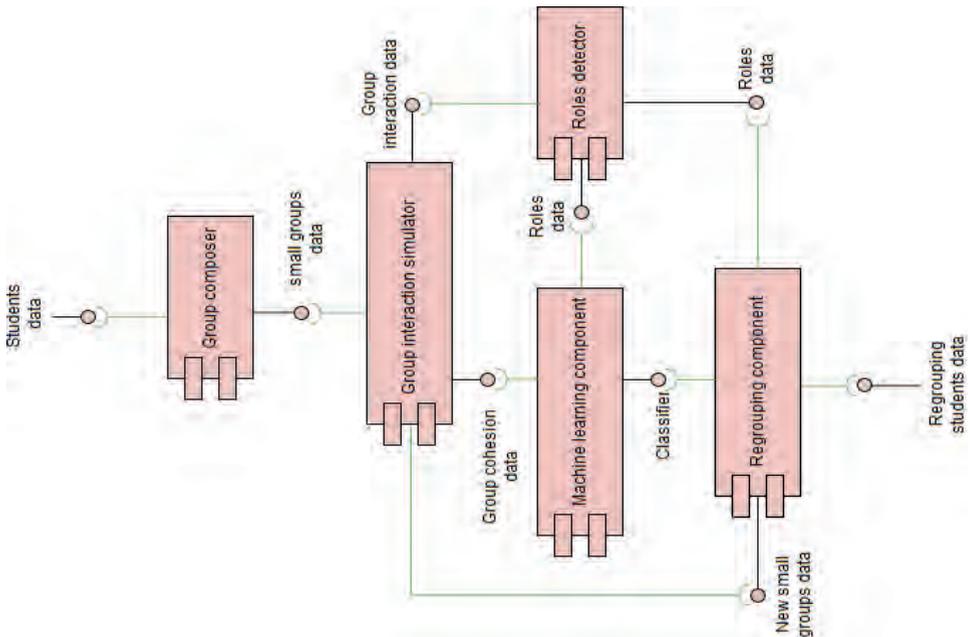


Fig. 2: Simulation system diagram

The Group composer implements the function of composing initial learning groups. It takes a set of students and group size as input. The output of this component is small learning groups of the given size. The initial groups can be composed at random. Note that in case of uneven split over all resulting groups (i.e. the last group could not have

the size of the given number), the last composed group either stay alone or merge into the last second group, depending on the actual size of the group and the assigned group size.

The Group interaction simulator generates the interaction data that is crucial for the next steps of the whole simulation. The heart of this component stems from Nygren's work on simulation of user participation and interaction in online discussion groups [Ny12]. Aiming at showcasing the social structure of group members, as is intuitively known to all, such an interaction simulator principally has to answer the following questions. 1) Who will start a discussion post at the next moment? 2) Will this post mention other participants (replies, comments and mentions of any sort)? 3) If so, who exactly will be mentioned? Regarding the first question, everyone has an opportunity to make a new post. The probability to make a new post is proportional to one's accumulative number of posts that have been already made. To answer the second question, Nygren used a probability parameter which observed from an empirical study to decide on the attachment of a groom (a groom could be comments and mentions of any sort). To answer the third question, they resorted to two strategies. First, the more often the participants are groomed, the more likely they have a chance to receive a groom again during the following interaction. Second, the more Groom-balance is, the more probability one has. Groom-balance is the difference between the number of grooms a participant has given away and the number of grooms he/she has received. Fig. 3 depicts the inner mechanism of the group interaction simulator.

```

parameters:
- P_nys //Probability to select a not-yet-spoken member as a new speaker
- P_groom_sum //probability to select a speaker according to the sum of grooms received
already
- P_groomed //probability to select a speaker out of groomed members:
- g_rate //the grooming rate
- P_groom_balance // probability to select the groomed person according to groom balance
- num_posts // the total number of posts

for (i=0; i<num_posts; i++):
//select a speaker (post maker) out of group members
if rand() > P_nys:
    Select a not-yet-spoken member
else: //select a speaker out of the yet-spoken members
    if rand() > P_groom_sum:
        Select the speaker with the max amount of grooms received
    else: //select a speak according to their grooming status
        if rand() > P_groomed:
            Randomly select one from the groomed group members
        else:
            select one from the ungroomed members
// Decide if a post contains a groom
if rand() > g_rate:
    Attach a groom to the post
//Decide whom is groomed if the post contains a groom
if rand() > P_groom_balance:
    Groom the member owning the max amount of groom balance
else:
    Groom the member made the max amount of posts

```

Fig. 3: Group interaction simulator

The Roles detector detects individuals' roles in each group. Roles identification has been studied for years. Social Network Analysis (SNA) has been recently used to

address this role identification problem [Su10, MMD15]. A critical question has to be answered when one applies SNA to detect group roles is how to map SN's metrics to specific roles. A wealth of studies has addressed this issue. Marcos et al. successfully identified isolated students, student-coordinators and teacher dependent students by observing SN's degrees, closeness and betweenness centrality [Ma08]. They recently extended it to detect more roles (e.g. teacher-facilitator) via applying their SNA tool, Role-AdaptIA [MMD15]. Rabbany et al. visualized the leaders and peripheral students using their Meerkat-ED tool [RTZ11]. Brokers as an important role in social networks were also studied by Stuetzer et al. using SNA and its SN characteristics have so far been uncovered [St13]. Referring to the aforementioned studies, the present work builds on six roles in total: *leader*, *disseminator*, *responder*, *broker*, *lurker* and *peripheral*. The mapping criteria can be seen in Tab. 1.

Roles	Degree	In-degree	Out-degree	Closeness-centrality	Betweenness-centrality
leader	H	M/H	M/H	H	
disseminator	M/H	M/H		M/H	
responder	M/H		M/H	M/H	
broker					M/H
lurker	N	N	N	N	N
peripheral	L	L/N	L/N	L	L/N

Note: H—High (bound: 0.7-1); M—Medium (bound: 0.2-0.7); L—Low (bound: 0-0.2); N—Null

Tab. 1: Criteria to Map SN Metrics into Group Roles

The Machine learning component trains a classifier that is able to predict the group outcome of any new form of group. Specifically, it sums up the number of students for each role in each group and regards group success as the target (in this paper, if the group cohesion of a group is larger than the median, it would be considered as a successful group). A classifier is a product of the machine learning component (i.e. a decision tree). It tells how good or bad the newly formed groups are. For example, a decision tree could indicate three group composition rules: 1) if the number of lurkers is not more than 2.5, the groups could be successful; 2) if the number of lurkers is more than 2.5 and the number of leaders is more than 1.5, the groups would then be good too; 3) if the number of lurkers is more than 2.5 and the number of leaders is not more than 1.5, we then would get unsuccessful groups. The total amount of predicted successful groups is a measure to evaluate the quality of the resulting group formations.

The Regrouping component functions as an optimizer. It creates new group formations for students who want to leave and iteratively optimizes grouping results with respect to the quality evaluated by the classifier. A discrete-PSO algorithm is applied to perform this optimization [ZP14].

3.3 Parameter settings

Regarding group interaction parameters, we almost borrowed all those parameter settings from the Nygren's work [Ny12], except for the following parameters. First of all, the number of posts sets up the total number of posts assumed to be made in each group in each task (70 in this work). The roles stereotyping rate is a probability to boost such active roles as leader and disseminator to make a new post. This is only applied to the tasks after the first one. Nygren did not observe the participants' behaviors in the subsequent tasks. The modeling is reported to be of Monte-Carlo type. If we run this modeling for the subsequent tasks the same as the first one, this would lead us to randomness. And this actually does not comply with our assumption that the group roles play an important part in group interaction. To avoid this, we applied an additional policy for regrouping that leaders and disseminators have some privileges to make a new post (determined by this stereotyping rate). It was set to 0.5 in this work.

Drop-out parameters define how many students drop out. Our simulator models the drop-out on a daily basis. Dropout students, by definition, are the most inactive students. Suppose a group task follows the pace of most MOOCs' weekly releasing mode, a weekly drop-out rate needs to be defined and the daily drop-out rate can just average it over 7 days of a week. Note that the vast drop out of MOOC students normally occurs in the first week (approx. 50%) followed by a comparatively smaller yet stable decline rate [K115, Ba13, MO13] in the following weeks. As such, placing a team task in the first week would not be a desirable thing to do because of the foreseeable group instability. Our parameters do not model the first week but assume a more or less stable decrease over time (as is realistic from week 2 on). From the second week onwards, we can estimate that the weekly drop-out rate could range from 0 to 50%. Still, we do not know how much exactly that weekly drop-out rate is. In practice, the answer should vary depending on different courses. With that in mind, there is no harm to set the weekly drop-out rate to a random float number ranging from 0 to 50%.

Regrouping parameters define that how many students from four different categories would leave the current groups for the newly composed ones. Categorizing students is based on their participation and group performance. A pair of active and inactive participation together with another pair of successful and unsuccessful group performance composes the four different categories in this set of parameters. The baseline of defining the leaving rates is threefold. First, active students should be more likely to leave for new groups than their inactive counterparts. Second, unsuccessful students should desire more chances to make a success via joining new groups. Hence, the active but unsuccessful students should more likely to leave than the others. Besides, the most conservative but 'most clever' could be the inactive but successful students. They, in fact, take full advantage of the group work without any substantial effort. They could continuously enjoy the benefits as a free rider so as not to leave the current groups unless any negative consequences are foreseeable. Based on the upper assumptions, we can simply randomize the leaving rates of those four categories of students (ranging from 0 to 1), but they should also follow the aforementioned three laws.

3.4 Simulation experiment

To examine the proposed group re-composition's impact via the simulator, two indicators were selected to be observed, namely group cohesion and drop-out. Note that we could not use learning performance as an indicator in our simulation (as no learning was modeled). Alternatively, group cohesion was selected as a substitute for the learning performance for the sake of their positive relationship to each other. Dating back to 1990s, Evans et al. found such positive relationship by the use of meta-analysis [ED91].

Group cohesion, as a structural measure of social network, in this work, is defined as the number of the actual inter-ties among group members divided by the total number of possible ties between any pair of members [Wi14]. By definition, group cohesion directly reflects inter-person ties in groups. The higher the group cohesion is, the stronger ties the groups have. Such strong ties, as inter-communication pipelines, undoubtedly address the salient problem of information-sharing and knowledge-sharing. Group members can thus better know each other and faster fulfil their common goals as a result [Ha99, LFF10]. Group cohesion, to some extent, mirrors group performance. It was thereby chosen as an indicator to learning performance in this simulation work.

Drop-out as another important indicator reflects students' engagement. In current MOOCs, students' engagement is reported to be associated with course content and personal motivations. In group work, what factors can explain such has never been studied thus far. The observation thereby, on the one hand, can hopefully imply some unseen hints to address the high drop-out problem. On the other hand, it can reveal the proposed group re-composition's impact.

Regarding the experimental procedure, first of all, we simulated those 10,000 students' (1,000 groups) group interactions in the first task. In the meantime, we regularly removed dropout students. After their interaction, we selected the students who had desired for a new group. The selection was based on the regrouping parameters mentioned in Section 3.3. We then composed them into new groups using the proposed approach. In order to highlight results in a comparable fashion, we also copied that number of students and composed them into random groups. The former is named algorithmic condition and the latter is named random condition. In the real world, certainly, this is not feasible. In simulation, it is however fairly easy and allows us to simulate what happens to the same students when re-grouped both algorithmically and randomly. We next simulated all those new groups' interactions in Task 2 and removed the dropout students again. When the second task was over, we counted the number of dropout students and the number of cohesive groups from both the algorithmic condition and the random condition. Note that a cohesive group is a group with a group cohesion that is higher than the median cohesion over all groups in both conditions. We ran the whole process 10 times in avoidance of any biased result probably generated by chance.

4 Results

4.1 Impact on group cohesion and dropout

Tab. 2 presents the group cohesion and drop-out. Recall that the whole simulation was repeatedly run 10 times intending to avoid by-chance results. As shown in the table, the amount of cohesive learning groups in the algorithmic condition is larger than in the random condition (algorithmic: 0.466 vs random: 0.421). A t-test was performed and the result indicates that the algorithmic condition produced significantly more cohesive groups than the random condition (p-value: 0.035), which shows its superiority on increasing group cohesion.

Runtimes	#Cohesive groups/total		#Drop-out/total	
	Algorithmic	Random	Algorithmic	Random
#1	0.447	0.335	0.428	0.571
#2	0.482	0.476	0.453	0.546
#3	0.420	0.411	0.445	0.554
#4	0.504	0.488	0.537	0.462
#5	0.440	0.36	0.503	0.496
#6	0.503	0.468	0.446	0.553
#7	0.486	0.464	0.465	0.534
#8	0.447	0.419	0.516	0.483
#9	0.480	0.360	0.416	0.583
#10	0.456	0.431	0.416	0.522
Ave.	0.466	0.421	0.468	0.530
SD	0.026	0.051	0.037	0.037
p-value	0.035		0.002	

Tab. 2: Number of cohesive Groups and Drop-out

Regarding drop-out, more dropout students came from the random groups than from the algorithmic ones (algorithmic: 0.468 vs random: 0.530). Likewise, a conducted t-test indicates a significant difference between both conditions (p-value: 0.002).

Although learning performance was not observable in this simulation experiment, it is not difficult to infer the positive impact on learning performance based on the reported positive ties between group cohesion and learning performance. Since we also saw a lower drop-out rate in the algorithmic groups, the impact on declining the drop-out rate appears positive too.

4.2 Impact of class size

Massive enrolment (thousands of participants) is one of the typical (and defining) MOOC features. Testing re-grouping methods on a data set of 10,000 students as we did is thus obviously necessary. However, one could ask whether the approach also works with a few thousand participants or with even smaller courses, such as an on-campus moodle course with a few hundred students typically. In an attempt to answer such a question, another two simulation experiments were run with three thousand students and one hundred students respectively.

When the participation is set to three thousand students, the observation on drop out and group cohesion reveals no difference to the case of ten thousand students. Similarly, more cohesive groups came from the algorithmic condition than the random one (average: 0.474 vs 0.428). Students tended to drop out more likely in the random condition (0.475 vs 0.524). A statistical test again confirmed significant differences (#cohesive groups: p-value = 0.011, dropout: p-value= 0.001).

In the case of one hundred students, the drop-out rate found in both conditions is almost same (average: 0.501 vs 0.498, p-value: 0.959). The reason for the deficits in this case has their roots in the very limited number of groups to be composed. The number of newly composed groups should be fewer than the total 10 groups (group size is set to 10). In such a small possibility scope, there is no need to challenge the algorithm's capability. In other words, the algorithmic method could perform no better than a random method in such a small case.

Varying the course size from one hundred to ten thousand over these three experiments, the observations tell us two points. First, the scale of participation does matter for the simulation results. Second, the proposed data-driven approach seems to make a positive effect beyond a certain level of participation. At least, for a hundred students, it does not reveal any of its superiorities.

5 Conclusions

This work proposed a novel data-driven approach to address the group re-composition problem that has not attracted much research attention thus far. The recent massive online open courses are arguably a fit to its application. A second contribution of the paper is the evaluation method via simulated classes. Through a simulation experiment, we saw its capability in decreasing the drop-out rate in groups and bringing more cohesive groups when compared against a random grouping strategy. Because of computer simulation's inherent drawbacks of stripping away realism to some extent, the findings are accordingly sensitive to the simulation settings and assumptions. We are certainly aware of this and therefore based the simulation on parameters and theories extracted from literature. Yet, certainly empirical evidence needs to be collected to back our findings.

Bibliography

- [Ba13] Balakrishnan, G.: Predicting student retention in massive open online courses using hidden markov models. Master's thesis, EECS Department, University of California, Berkeley, 2013.
- [ED91] Evans, C.R.; Dion, K.L.: Group Cohesion and Performance: A Meta-Analysis. *Small Group Research* 22, 175-186, 1991.
- [Ga86] Gamoran, A.: Instructional and Institutional Effects of Ability Grouping. *Sociology of Education* 59, 185-198, 1986.
- [GB06] Graf, S.; Bekele, R.: Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In: *Proc. Int. Conf. on Intelligent Tutoring Systems*, Springer-Verlag, 2171765, S. 217-226, 2006.
- [Ha99] Hansen, M.T.: The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly* 44, S. 82-111, 1999.
- [Hs14] Hsu, C.-C.; Chen, H.-C.; Huang, K.-H.; Huang, K.-K.; Huang, Y.-M.: The development of an adaptive group composition system on Facebook for collaborative learning using an artificial bee colony algorithm. *Appl. Math* 8, S. 157-164, 2014.
- [Jo14] Jordan, K.: MOOC Completion Rates: The Data. <http://www.katyjordan.com/MOOCproject.html>, Stand: 27.08.2014.
- [KI15] Kloft, M.; Stiehler, F.; Zheng, Z.; Pinkward, N.: Predicting MOOC Dropout over Weeks Using Machine Learning Methods. In: *Proc. of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses*, S. 60-65, 2015.
- [KBS14] Konert, J.; Burlak, D.; Steinmetz, R.: The Group Formation Problem: An Algorithmic Approach to Learning Group Formation. *Open Learning and Teaching in Educational Communities*, S. 221-234. Springer, 2014.
- [Ku92] Kulik, J.A.: *An Analysis of the Research on Ability Grouping: Historical and Contemporary Perspectives*. Research-Based Decision Making Series, 1992.
- [LFF10] Lechner, C.; Frankenberger, K.; Floyd, S.W.: Task Contingencies in the Curvilinear Relationships Between Intergroup Networks and Initiative Performance. *Academy of Management Journal* 53, S. 865-889, 2010.
- [Ma08] Marcos-Garcia, J.A.; Martinez-Monés, A.; Dimitriadis, Y.; Rodriguez-Triana, M.J.: *Role-AdaptIA: A role-based adaptive tool for interaction analysis*, 2008.
- [MMD15] Marcos-García, J.-A.; Martínez-Monés, A.; Dimitriadis, Y.: DESPRO: A method based on roles to provide collaboration analysis support adapted to the participants in CSCL situations. *Comput Educ* 82, S. 335-353, 2015.
- [MO13] MOOCs@Edinburgh-Group: MOOCs @ Edinburgh 2013 – Report #1, 2013.

- [MOV12] Moreno, J.; Ovalle, D.A.; Vicari, R.M.: A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Comput Educ* 58, S. 560-569, 2012.
- [Ny10] Nygren, E.: Modeling the social dynamics of online discussion sites. In: *Proc. Int. Workshop on Modeling Social Media*, ACM, 1835982, S. 1-8, 2010.
- [Ny11] Nygren, E.: Grooming Analysis Modeling the Social Interactions of Online Discussion Groups. In (Atzmueller, M.; Hotho, A.; Strohmaier, M.; Chin, A., Hrsg.): *Int. Workshops on Analysis of Social Media and Ubiquitous Data*, S. 37-56, 2011.
- [Ny12] Nygren, E.: Simulation of User Participation and Interaction in Online Discussion Groups. In (Atzmueller, M.; Chin, A.; Helic, D.; Hotho, A., Hrsg.): *Int. Workshops on Modeling and Mining Ubiquitous Social Media*, S. 138-157, 2012.
- [RTZ11] Rabbany, R.; Takaffoli, M.; Zai'ane, O.R.: Analyzing participation of students in online courses using social network analysis techniques. In: *Proc. Int. Conf. on Educational Data Mining*, 2011.
- [SA10] Shimazoe, J.; Aldrich, H.: Group Work Can Be Gratifying: Understanding & Overcoming Resistance to Cooperative Learning. *College Teaching* 58, S. 52-57, 2010.
- [SB14] Srba, I.; Bielikova, M.: Dynamic Group Formation as an Approach to Collaborative Learning Support. *Learning Technologies*, IEEE Transactions on, 2014.
- [SI87] Slavin, R.E.: Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis. *Review of Educational Research* 57, S. 293-336, 1987.
- [St13] Stuetzer, C.M.; Koehler, T.; Carley, K.M.; Thiem, G.: "Brokering" Behavior in Collaborative Learning Systems. *Procedia - Social and Behavioral Sciences* 100, S. 94-107, 2013.
- [Su10] Sundararajan, B.: Emergence of the most knowledgeable other (MKO): Social network analysis of chat and bulletin board conversations in a CSCL system. *Electronic Journal of e-Learning* 8, S. 191-208, 2010.
- [We16] Wen, M.; Maki, K.; Wang, X.; Dow, S.P.; Herbsleb, J.; Rose, C.: Transactivity as a Predictor of Future Collaborative Knowledge Integration in Team-Based Learning in Online Courses. In: *Proc. Int. Conf. on Educational Data Mining*, 2016.
- [Wi14] Wise, S.: Can a team have too much cohesion? The dark side to network density. *European Management Journal* 32, S. 703-711, 2014.
- [YA12] Yannibelli, V.; Amandi, A.: A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. *Expert Systems with Applications* 39, S. 8584-8592, 2012.
- [ZP14] Zheng, Z.; Pinkwart, N.: A Discrete Particle Swarm Optimization Approach to Compose Heterogeneous Learning Groups. In: *Proc. Int. Conf. on Advanced Learning Technologies (ICALT)*, S. 49-51, 2014.

Teamwork assessment and peerwise scoring: Combining process and product assessment

Ian G. Kennedy¹, Paul H. Vossen²

Abstract: Teamwork is commonly required in the industry [Sa95]. At university, students learn teamwork by working in groups, for example during a software development assignment. Teamwork affects their productivity. Lecturers evaluate their overall group performance. Lecturers also need to assess the cooperation of each team member. Cooperation data comes from the students involved, e.g., via a questionnaire. Here we draw attention to specific items for assessment. We then show how to merge or aggregate evaluations using the lecturer's overall quality scale and the students' peer assessment scale. To this end, we will explain and demonstrate the underlying Split-Join Invariance principle using two compatible scoring formulae.

Keywords: Grading, Peer Assessment, Performance Rating, Quality Scale, Scoring Rule, Scoring Rubric, Split-Join Invariance (SJI), Team Cooperation

1 Design of assessment for teamwork

Think of this realistic scenario. A lecturer has a software engineering class of 35 students. They have to work in 8 groups of 3 or 4 students. There is no final exam, e.g. a multiple-choice test, just the outcome of teamwork. Scores are to be reported on an individual basis, though, and the amount of assessment will be about a half hour per team. How to assess doing justice to all involved? Students who are grade grubbing manipulate overall impression marks easily, and there may be free riders as well. A fair and robust assessment will not be challenged by students. A fair assessment will use attributes which are made known beforehand, along with the rule for combining the attributes, and any weights thereof. The specified rules must be consistently applied to all students and groups, and be traceable. Moreover, arbitrariness, discrimination or dishonesty should be avoided by awarding the marks mechanistically, according to publicised norms. Finally, the marks should not pretend to be more accurate than they are; their production can be replicated.

In designing courses, e.g., practice-oriented software engineering courses, we consider challenges involved in preparing students for the real world. After graduation, students will be required to work in teams, e.g., to build large software systems [Sa95]. Thus, they

¹ Ex University of the Witwatersrand, Johannesburg, South Africa, 2000, dr.iankennedy@gmail.com

² Research Institute SQUIRE, Kiefernweg 1A, Niederstetten, D-97996, Germany, p.h.vossen@googlemail.com

must now learn to work in teams. A textbook knowledge of teamwork assessed by a written exam on it is clearly insufficient: *students learn teamwork only by doing it*. Students need to reflect and evaluate their team cooperation in a self-critical way. Lecturers –in addition to their assessment of the overall team deliverable– need detailed and reliable information about the teamwork. Such data can only come from the students, e.g., through questionnaires or by automated measurements during or after group work. Our questions are: What data is required for appropriate, meaningful peer and team assessment? How can the results of the peer assessment be combined with the assessment of the productivity and products of the group work into a final evaluation of each individual student?

2 Attributes in the PROCESS scoring rubric

After finding only inferior references, we invented our own list of attributes of cooperation (Table 1, based on [VK17]) that we feel all team members should learn. These attributes can be revised, localised and then taught during introductory lectures. They can be tested and assessed via a group project. Here are our proposed attributes (scoring rubrics):

Outcome no.	PROCESS Learning Objective
1	Commitment
2	Collaboration
3	Coaching each other
4	Coordination
5	Control
6	Communication
7	Climate fostering
8	Constructive criticism
9	Coping with adversity and diversity
10	Conflict management

Tab. 1: Specimen individual PROCESS rubric

Commitment: It is easy to overlook the importance of commitment. Team members must see the project goal or the corporate goal and remember that they pass, or are only paid a salary to produce software successfully and align their allegiance to this aim.

Collaboration: Collaboration is working jointly towards the agreed-upon project goal. How well does each team member facilitate the work of others in the team?

Coaching each other: Team members must learn from each other and teach each other.

Coordination: Software team members have to coordinate their activities to ensure, e.g., that they do not save their work on top of others' work.

Control: An airline pilot takes the final responsibility for the control of the aeroplane. Although the co-pilot can fly the plane, only the pilot has a hand on the joystick in the cockpit. So too, a chain of command in the software team will develop.

Communication: Communication includes listening, and written as well as spoken com-

munication. It includes upward, downward and horizontal communication inside and outside the group.

Climate fostering: Team members should help each other to maintain an open and positive team climate for discussions.

Constructive criticism: It is important to help others with constructive criticism.

Coping with adversity and diversity: Coping with difficult circumstances includes, e.g., coping with absenteeism and change requests.

Conflict management: In practice, there always will be conflict in a team. The question is: how well does the team member overcome the conflicts and resolve the issues?

3 Attributes in the PRODUCT scoring rubrics

Table 2 shows our specimen rubric for software development [VK17]. We do not propose a final or standard list of assessment criteria. We include the specimen here to make our paper more readable and practice-oriented. The list of PRODUCT outcomes in Table 2 is just an example because it is highly dependent on the application area as to which attributes will be assessed. The PRODUCT is a software application including documentation. Each item in the table has a deliverable attached to it, which the lecturer will evaluate.

Outcome no.	PRODUCT Learning Objective
1	Purpose of project
2	Project specification
3	Library review
4	Functionality
5	Usability
6	Portability
7	Maintainability
8	Documentation
9	Timeliness
10	References

Tab. 2: Specimen group PRODUCT rubric

4 Technical aspects

If the lecturer uses peer assessment only for formative evaluation, then there is no need to go beyond simple questionnaire-based scoring rubrics. In practice, however, peer assessment is frequently used to produce individual student scores, i.e., for summative evaluation. The peer assessment procedure outlined above can inform the lecturer about team dynamics. Such data evaluate the individual student's cooperation. Appropriate rating methods, e.g. Likert scales, including possible weighting of assessment criteria or rubrics can be used to turn the captured data into quantitative statements. In the next section, we examine the necessary and sufficient conditions for such scoring formulae to yield a sound, robust and fair overall evaluation of each student involved in a team.

4.1 Split-Join-Invariance (SJI)

The most important principle for consistent, robust and fair peer assessment is called Split-Join-Invariance. See also [Vo11, Vo14, Vo17, VK17]. If we split the overall team score into single student scores, and then join those student scores by a suitable aggregation or averaging function, e.g. the arithmetic or geometric mean, the result must be equal to the initial team score given by the lecturer. *Older models for summative peer assessment missed or ignored this simple but powerful requirement.* There are two reasons for requiring SJI. The first is a conceptual reason, and can be explained by simply asking: *what else could we mean with overall team score as an aggregated or averaged evaluation of the joint work of all team members?* The point is that there is no *a priori* guarantee that the calculated average of student scores will equal the lecturer's initial team score. We have to build this into our scoring formula. The second reason is equally important. Let us ask again a simple question: *what could happen if we were not careful in our choice of a scoring formula so that the mean student score was not equal to the overall team score from which the student scores were derived?* If the mean student score is lower than the team score, then some students may rightly complain that they do not get full credit for their involvement in the joint work. In the case of a higher mean student score, the students might demand that their scores be recalculated based on this higher mean score. So, for consistency, plausibility, acceptability and accountability reasons, every peer assessment scoring formula – based on the arithmetic or geometric or any other plausible mean – must satisfy the SJI principle.

4.2 SJI-incompatible versus SJI-compatible approaches

Our [VK17] extensive literature research covering more than 25 years of publications revealed that no scoring formula proposed for peer assessment took SJI or a similar principle into account. Instead, scoring rules were proposed based on their *intuitive* plausibility or simplicity, not for reasons of logical or mathematical *appropriateness* like correctness, consistency, robustness and fairness. Previous peer-assessment scoring formulae fell into the category of linear functions in two variables: a *team score* and a *student rating*. Authors dealt mainly with the question: how can the two values (a rating and a score) be combined into a single value (score) by adding or multiplying, with or without weighting? We know of only one proposal [Ne12] which goes beyond this simple framework and adds a quadratic (or *parabolic*, in the terminology of the author) component to prevent individual scores from growing too large. All such solutions (except [Ne12]) are flawed in that final student scores can transcend the boundaries of the scoring scale. In all, the average student score need not equal the initial team score set by the lecturer, violating our Split-Join-Invariance principle [Vo17].

There are two main classes of SJI-compatible scoring rules. One class adopts the arithmetic mean (AM) and the other one the geometric mean (GM) for averaging student scores. Whereas there is only one generalised scoring scheme associated with GM, there are (at least) two distinct scoring schemes within the class of AM-based scoring rules: symmetric

versus asymmetric scoring formulae. Each class or subclass has a single scoring rule, i.e., a non-linear scoring function with one or two parameters. For space reasons, we will only sketch the two general scoring formulae for the class of arithmetic scoring functions. Let us first introduce the terminology and symbols to be used. The team score based on lecturer's assessment of the team's final PRODUCT is t . A student's PROCESS rating based on peer assessments by all peers in his or her team is σ_i for student i . It is a number between $-k$ and $+k$ on a bipolar Likert scale (a generalisation to the entire real line will appear in a forthcoming paper). The degree to which this student rating is to have an impact on the team score will be denoted by $\varphi(t,r)$, $\varphi_{i,r}$ or just φ . Here, t is the team score as introduced above; r is an integer or real number indicating faculty's or lecturer's tolerance or intolerance regarding the peer assessment impact. For instance, in the AM-case $r = 0$ corresponds to a medium tolerance level. Finally, there is a unique formula for each class or subclass which combines all these ingredients to produce a SJI-compatible student score s_i within the range of scores, i.e., between 0 and 1. The key ingredient is the expression for $\varphi_{i,r}$, which is different for symmetric or asymmetric AM-based scoring functions.

4.3 Symmetric scoring formula

Our [VK17] symmetric AM-based scoring formula has the following form, where τ is defined on the bipolar Likert rating scale $-k \dots +k$ and denotes the overall team rating:

$$s_i = \begin{cases} t + t \times \frac{1}{1+t^r} \times \frac{\sigma_i - \tau}{k + \tau}, & \sigma_i < \tau \\ t + (1 - t) \times \frac{1}{1+t^r} \times \frac{\sigma_i - \tau}{k - \tau}, & \sigma_i \geq \tau \end{cases} \quad (1)$$

Score s_i can never escape the valid range of scores [0,1]. The tolerance parameter r can be any positive or negative real number. Taking $r = 0$ gives a φ of 0.5, representing a mediocre tolerance level. Positive r means a high(er) tolerance $\varphi > 0.5$; negative r means a low(er) tolerance $\varphi < 0.5$. In practice, integer values may be taken for r , e.g. values between -10 and +10. Taking $r = 1$ or 2 gives easy to handle, tolerant φ 's. Taking $r = -1$ or -2 is suitable for the less tolerant mindset. It is possible to transform (1) into special, simpler looking formulae if required. What about τ ? It should be such that the SJI principle holds. It turns out that τ is equal to the ψ -arithmetic mean of student rates σ_i for ψ equal to the scoring function (1).

4.4 Asymmetric scoring formula

Our [VK17] general asymmetric AM-based scoring formula has the following structure, where τ is the same as above and tolerance $r \geq 1$:

$$s_i = \begin{cases} t + t \times (1 - t^{r-1}) \times \frac{\sigma_i - \tau}{k + \tau}, & \sigma_i < \tau \\ t + (1 - t) \times (1 - (1 - t)^{r-1}) \times \frac{\sigma_i - \tau}{k - \tau}, & \sigma_i \geq \tau \end{cases} \quad (2)$$

The most important point about the asymmetric arithmetic scoring formula is that low student ratings ($\sigma_i < \tau$) are handled differently from high student ratings ($\sigma_i > \tau$). If we increase r , the tolerance is increased and vice versa. For $r = 1$, we effectively prohibit peer ratings from having any influence on the student score. An exceptional case occurs for $r = 2$: it is the only case in which lecturer's team score falls within a range of tolerance which has the same length $t \times (1-t)$ left of t as right from t . We call this unique case the *balanced tolerance range scoring formula* (not to be confused with a symmetric scoring formula). Again, it turns out that τ equals the ψ -arithmetic mean of student rates σ_i for ψ equal to the scoring function (2).

5 Discussion and Conclusion

This paper has raised awareness about the categories and items that need to be assessed for teamwork results and showed how we could get student ratings for the cooperation, e.g., in a software team, based on peer ratings by other members of the team. It has also given a scoring rubric for the lecturer to mark the group. Finally, the work has shown how to consistently, robustly and fairly combine these into an aggregated evaluation for the student's reports, all based on the important Split-Join-Invariance principle. We also provided some example scoring functions.

References

- [Ne12] Nepal, K. P.: An approach to assign individual marks from a team mark: the case of Australian grading system at universities. *Assessment & Evaluation in Higher Education*, 37/5, S. 555-562, 2012.
- [Sa95] Saeki, M.: Communication, collaboration and cooperation in software development: how should we support group work in software development? In: *IEEE Software Engineering Conference, Asia Pacific*, S. 12-20, 1995.
- [VK17] Vossen, P.; Kennedy, I.: A fair group marking scheme combining process and product assessment, Forthcoming in *Practitioner Research in Higher Education Conference*, University of Cumbria, Manchester, S. 14, 2017.
- [Vo11] Vossen, P. H.: A truly generic performance assessment scoring system. In (Gómez Chova, L.; Candel Torres, I.; López Martínez, A., Hrsg.): *Proc. of the INTED 2011 Conference*, Valencia - Spain, March 7-9, S. 2448-2458, 2011.
- [Vo14] Vossen, P. H.: Educational Assessment Engineering: A Pattern Approach. In (Balas, V.E.; Jain L.C.; Kovačević B., Hrsg.): *Soft Computing Applications*, Springer International Publishing, S. 605-619, 2014.
- [Vo17] Vossen, P. H.: Distributive Fairness in Educational Assessment: Psychometric Theory meets Fuzzy Logic. In (Balas et. al., Hrsg.): *Proceedings of the SOFA 2016 Conference*. Springer Advances in Intelligent Systems and Computing, Springer International Publishing, S. 17, 2017 (in press).

StudiDuell App – Mobiles Lernen mit interaktiven Hörsaalspielen

Sebastian Hobert¹, Almut Reiners², Pascal Freier¹, Matthias Schumann¹

Abstract: Hörsaalspiele können in der universitären Hochschullehre eingesetzt werden, um Studierende durch spielerische Elemente (sog. Gamification) zu motivieren, sich mit Lerninhalten auseinander zu setzen. Ein ähnlicher motivationssteigernder Effekt, der zu einer steigenden Lerner-Content-Interaktion führt, ist durch den sinnvollen Einsatz von Smartphones und Tablets (sog. Mobiles Lernen) in der Lehre sichtbar. Ziel des im Rahmen dieses Beitrags entwickelten Prototyps „StudiDuell“ ist es daher, die Konzepte *mobiles Lernen* und *Hörsaalspiele* zu kombinieren, um eine mobile Anwendung für Smartphones bereitzustellen, mit der Studierende auch in großen Lehrveranstaltungen an interaktiven Hörsaalspielen teilnehmen können. Dazu stellt der Beitrag das zugrundeliegende didaktische Konzept des Hörsaalspiels vor, das darauf aufbauende technische Konzept sowie den entwickelten Prototypen der Smartphone-App. Als Ergebnis wird schließlich die Eignung der entwickelten App für den Einsatz in der universitären Hochschullehre durch eine in einer Lehrveranstaltung mit 30 Bachelorstudierenden durchgeführten Evaluation nachgewiesen.

Keywords: Mobiles Lernen, Hörsaalspiel, Interaktivität, Lehre, Gamification

1 Einleitung

Der Einsatz von spielerischen Elementen in spielfremden Kontexten (sog. Gamification; [De11]) ist aus vielen Anwendungen (insbesondere Smartphone-Apps) nicht mehr wegzudenken. Damit wird das Ziel verfolgt, die Motivation der Nutzerinnen und Nutzer zu steigern und dadurch die durchschnittliche Nutzungsdauer der Anwendungen zu verlängern [Sa13]. Typische Spielelemente, die sich in unzähligen Anwendungen wiederfinden lassen, sind Badges, automatisiertes Feedback oder Ranglisten, um den spielerischen Wettbewerb zwischen Nutzerinnen und Nutzern anzuregen [De11].

Der motivationsfördernde und aktivierende Charakter von Gamification wurde auch in Lernanwendungen bereits als positiv für den Lernerfolg nachgewiesen [SG05], [Ho15], [Ha14]. Insbesondere im Bereich des mobilen Lernens haben Studien gezeigt, dass das Potenzial besteht, die Interaktion zwischen Lernenden und den zu vermittelnden Lerninhalten (sog. Lerner-Content-Interaktion) zu steigern [Ha05]. Auch wenn der Einsatz von innovativen Technologien in Kombination mit spielerischen Elementen den

¹ Universität Göttingen, Professur für Anwendungssysteme und E-Business, Platz der Göttinger Sieben 5, 37073 Göttingen, {shobert,pfreier,mschuma1}@uni-goettingen.de

² Universität Göttingen, E-Learning-Service, Platz der Göttinger Sieben 5, 37073 Göttingen, almut.reiners@uni-goettingen.de

Lernerfolg positiv beeinflussen kann, werden solche Anwendungen in der universitären Präsenzlehre bisher nur selten eingesetzt. Eine Möglichkeit, mit der spielerische Elemente Einzug in Vorlesungen und Seminare finden können, sind Hörsaalspiele. Diese verbinden das Vermitteln, Festigen und Überprüfen von Lerninhalten mit spielerischen Elementen und sind mit variablen Gruppengrößen in Hörsälen und Seminarräumen spielbar [LSS14]. Auch wenn Hörsaalspiele ohne den Einsatz von digitalen Technologien möglich sind [LSS14], bieten insbesondere Smartphones und Tablets die Möglichkeit, Hörsaalspiele flexibel durchzuführen und eine größere Beteiligung seitens der Studierenden zu erreichen.

Ziel des vorliegenden Praxisbeitrags ist es daher, die Eignung des Einsatzes einer Smartphone-App für interaktive Hörsaalspiele im formalen Bildungskontext der universitären Lehre zu untersuchen. Dazu wird im Folgenden das zugrundeliegende didaktische Konzept des Hörsaalspiels StudiDuell, dessen technische Umsetzung sowie die Ergebnisse einer durchgeführten Evaluation vorgestellt.

2 Didaktisches Konzept

StudiDuell wurde entwickelt, um in lehrerzentrierten Vorlesungen oder Seminaren Studierende zu aktivieren, zur Auseinandersetzung mit Lerninhalten zu motivieren und die Interaktion zwischen Lehrenden und Studierenden zu verbessern. Dafür wurde eine Spielidee ausgewählt, die vor allem für die spielerische Übung, Wiederholung oder Festigung von Lerninhalten geeignet ist. Ziel des Spiels ist es, mehrere Gruppen von Studierenden im Team gegeneinander Quizfragen beantworten zu lassen. Durch direkte Rückmeldung zu den beantworteten Fragen sowie dem Einsatz einer Rangliste soll der kompetitive Ehrgeiz der Studierenden gesteigert werden.

Zum Durchführen des Spielkonzepts bereiten Dozierende das Spielfeld bestehend aus Themenfeldern sowie Punktzahlen vor. Weiterhin werden für jedes Themenfeld und jede Punktzahl geeignete Quizfragen (z. B. Single-Choice, Multiple-Choice und Freitext-Frage) benötigt. Lehrende entscheiden zudem, in Abhängigkeit der Anzahl der beteiligten Studierenden, in wie vielen Gruppen die Quizfragen gespielt und wie die Antworten in der Gruppe abgegeben und gezählt werden sollen. Möglich ist, dass alle Gruppenmitglieder antworten und die mehrheitsfähige Antwort gezählt wird (Beteiligung aller Studierenden), oder dass nur ein Sprecher der Gruppe antwortet (Förderung der Interaktion zwischen Studierenden). Die Lehrenden starten das Spiel, indem sie von einer Gruppe eine Frage auswählen lassen. Für das korrekte Beantworten einer Frage erhält die Gruppe die jeweilige Punktzahl. Beantwortet eine Gruppe die Frage falsch, kann die Frage an die nächste Gruppe weitergegeben werden. Gewonnen hat die Gruppe mit den meisten Punkten. Je nach vorhandenem Zeitfenster kann das Spiel durch Anpassung der Fragenmenge länger oder kompakter gestaltet werden.

Die Dozierenden haben mit diesem Spielansatz die Möglichkeit, ein direktes Feedback über den aktuellen Wissenstand der Studierenden zu erhalten. Während des Spiels

können Dozierende richtige und falsche Antworten mit den Studierenden in der Veranstaltung diskutieren und somit Wissenslücken schließen. Studierende erhalten durch die Teilnahme am Spiel die Möglichkeit, direktes Feedback zu ihrem aktuellen Kenntnisstand zu erhalten. Alternativ gibt es weitere didaktische Einsatzmöglichkeiten: So kann das Spiel zu Beginn einer Themeneinheit zur Aktivierung von Vorwissen genutzt werden, oder im Sinne des „Jeopardy-Prinzips“ umgestaltet werden, indem statt Fragen, Antworten vorgegeben werden können und die passende Frage gefunden werden muss. Weiterhin ist auch ein eher studierendenzentrierter Ansatz denkbar, bei dem Studierende zur Vorbereitung einer Vorlesung für einzelne Kategorien Fragen generieren und vor der Veranstaltung einreichen, um diese anschließend im Rahmen eines StudiDuells zu diskutieren [Lu14].

3 Konzeption und Implementierung

Um das zuvor beschriebene Spielkonzept im Hörsaal umsetzen zu können, soll die Lernanwendung sowohl die Dozierenden beim Steuern und Verwalten des Hörsaalspiels, als auch die Studierenden bei der Beantwortung der Quizfragen unterstützen. Dazu wurde ein aus zwei Komponenten bestehendes Konzept erarbeitet (siehe Abbildung 1): Einerseits wird eine mobile Anwendung – die StudiDuell-App – benötigt, mit der Studierende Zugriff auf die jeweiligen Quizfragen bekommen sowie ihre Antwort eingeben können. Andererseits wird eine zweite Komponente benötigt, mit der Dozierende das Spiel verwalten, steuern und über einen Beamer im Hörsaal präsentieren können.

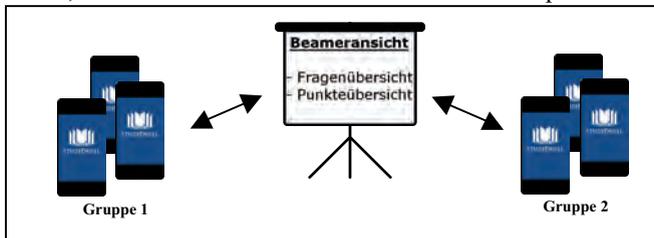


Abbildung 1 Umsetzungskonzept von StudiDuell

Der Nutzungsablauf des StudiDuell-Konzepts sieht dabei vor, dass Dozierende das Hörsaalspiel über die Desktop-Anwendung starten und den für das Spiel erzeugten QR-Code anzeigen. Nachdem die Studierenden die StudiDuell-App über das Einscannen des QR-Codes auf ihren Smartphones gestartet haben, werden sie in mindestens zwei Gruppen eingeteilt³. Anschließend wählt die erste Gruppe ihre Fragenkategorie aus und alle Studierenden der ersten Gruppe haben die Möglichkeit, die Frage auf den eigenen Smartphones zu beantworten (siehe Screenshots in Abbildung 2). Falls die Mehrheit der Gruppe sich für die richtige Antwort entschieden hat, erhält sie die dafür vorgesehene Punktzahl. Andernfalls darf die nächste Gruppe die gleiche Frage beantworten („Frage

³ Die StudiDuell-App bietet den Dozierenden die Möglichkeit, die Anzahl der Gruppen variabel zu definieren, um bspw. das Hörsaalspiel auch in Massenveranstaltungen mit vielen Gruppen einzusetzen.

schieben“) oder aus den verbleibenden Kategorien eine neue Frage auswählen. Nachdem alle Fragen durch die Studierendengruppen nacheinander beantwortet wurden, wird die finale Punkteübersicht über die Desktop-Anwendung präsentiert.



Abbildung 2 Screenshots der StudiDuell App

Die Entwicklung beider Komponenten erfolgte mit Webtechnologien (u. a. mit HTML5 unter Verwendung des Bootstrap-Frameworks auf Clientseite, sowie PHP und MySQL auf Serverseite), um eine flexible Nutzung mit Smartphones, Tablets und Desktop-Computern zu ermöglichen. Durch die Entwicklung einer webbasierten App wurde insbesondere die Problematik der Heterogenität vorhandener Smartphone-Betriebssysteme (z. B. Android, iOS oder Windows Phone) umgangen, da lediglich ein Webbrowser zum Zugriff auf die StudiDuell-App benötigt wird [C111]. Auf diese Weise wird es jedem Studierenden mit einem internetfähigen Endgerät ermöglicht, an StudiDuell-Hörsaalspielen teilzunehmen.

4 Evaluation

Zur Evaluation der mobilen Anwendung wurde diese im Rahmen einer Lehrveranstaltung an der Universität Göttingen eingesetzt und anschließend mittels quantitativem Fragebogen mit Studierenden ($n=30$) evaluiert. Bei den Befragten handelte es sich um Bachelorstudierende der Wirtschaftsinformatik zwischen dem dritten und achten Fachsemester im Alter von 21 bis 24 Jahren. Der Gestaltung des Fragebogens wurde dabei das Technology Acceptance Model (TAM; [DBW89]) zugrunde gelegt, sodass die Nützlichkeit, Einfachheit sowie Nutzungseinstellung und -absicht von StudiDuell über hinterlegte Konstrukte untersucht wurde. Die Messung der abgefragten Konstrukte erfolgte anhand einer fünfstufigen Likert-Skala (-2: trifft nicht zu; +2: trifft zu).

Die Auswertung ergab, dass die Befragten sämtliche abgefragten Konstrukte des TAM-

Models als positiv bewerten. Insbesondere stellten die Studierenden heraus, dass das Spielkonzept (Mittelwert: +1,9; Median: +2,0) sowie die Bedienung der App (Mittelwert: +1,8; Median: +2,0) einfach zu verstehen waren. Dies zeigt sich auch darin, dass die Darstellung der App auf den Smartphones (Mittelwert: +1,8; Median: +2,0) sowie dem Beamer (Mittelwert: +1,7; Median: +2,0) übersichtlich war und dass die nächsten Schritte jederzeit nachvollziehbar waren. Lediglich die durch das Bilden von konkurrierenden Teams entstehende Notwendigkeit von Diskussionen und dem Austausch zwischen den Studierenden innerhalb eines Teams (sog. Lerner-Lerner-Interaktion) wurde von den Probanden nur mit durchschnittlich gut empfunden, sodass ein Großteil der Studierenden die Fragen eher eigenständig ohne eine Gruppe beantworten würden (Mittelwert: +0,3; Median: +0,5). Gerade dies zeigt, das Training für Gruppenarbeiten durchaus sinnvoll erscheint. Gleichwohl ergab die Auswertung, dass die Konkurrenz zwischen den Studierenden bzw. Gruppen als motivierender Faktor dafür angesehen werden kann, sich mit den Lehrinhalten zu befassen (Mittelwert: +0,6; Median: +1,0). Zudem bereitete den Studierenden das Hörsaalspiel Spaß (Mittelwert: +1,3; Median: +2,0). Somit wirkt sich das gewählte Spielprinzip und der Gamification-Ansatz positiv auf die Lerner-Content-Interaktion aus.

In einer abschließenden Gesamtbewertung gaben die Probanden an, die StudiDuell-App insgesamt als sinnvoll anzusehen (Mittelwert: +1,6; Median: +2,0). Ein Einsatz von mobilen Hörsaalspielen in der universitären Lehre empfinden sie dementsprechend als Mehrwert. Dies zeigt sich auch darin, dass die Studierenden Lehrveranstaltungen, die Hörsaalspiele wie StudiDuell einsetzen, ihren Kommilitonen weiterempfehlen würden (Mittelwert: +1,4; Median: +2,0). Abbildung 3 stellt beispielhaft vier Boxplots der analysierten Konstrukte dar.

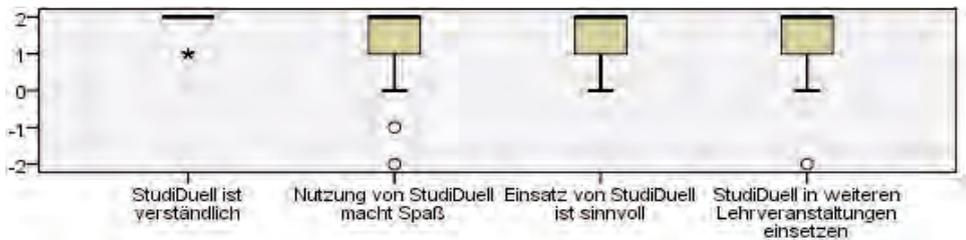


Abbildung 3 Ausgewählte Boxplots der analysierten Items

5 Ausblick

Nach der positiven Evaluation der StudiDuell App sowie dem Einsatz in einer Lehrveranstaltung wird die StudiDuell App zukünftig für Dozierende der Universität für den Einsatz in Lehrveranstaltungen zur Verfügung stehen. Dadurch soll das Spektrum an verfügbaren mobilen Lernangeboten erweitert werden, sodass ein zusätzliches Mittel zur Motivation und Aktivierung von Studierenden zur Verfügung gestellt wird. Neben dem direkten Beitrag für die Praxis, kann der vorliegende Artikel auch als Grundlage für weitere Forschungsarbeiten dienen. Aufbauend auf den Ergebnissen gilt es bspw. zu

untersuchen, welche weiteren Formen interaktiver Hörsaalspiele sich für den Einsatz mit großen Gruppen in Lehrveranstaltungen eignen. Dazu sollte einerseits der didaktische Hintergrund beleuchtet werden, andererseits sollten mögliche Umsetzungskonzepte in Massenveranstaltungen evaluiert werden, um optimale Rahmenbedingungen und Voraussetzungen ermitteln zu können.

Danksagung

Wir danken den Studierenden Christian Finke, Michael Groth, Julian Meyer, Naresh Visuvalingam, Maximilian Flock und Dennis Scholl für die prototypische Implementierung der StudiDuell App im Rahmen des *Projektseminars Systementwicklung – Entwicklung mobiler Anwendungen* im Sommersemester 2016.

Literaturverzeichnis

- [Cl11] Clevenger, N.: iPad in the Enterprise: Developing and Deploying Business Applications. John Wiley & Sons, Indianapolis, IN, 2011.
- [DBW89] Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R.: User Acceptance of Computer Technology. A Comparison of Two Theoretical Models. *Management Science* 8, S. 982–1003, 1989.
- [De11] Deterding, S. et al.: From game design elements to gamefulness. In (Lugmayr, A. et al., Hrsg.): Proceedings of the 15th International Academic MindTrek Conference Envisioning Future Media Environments. ACM, New York, NY, S. 9–15, 2011.
- [Ha05] Hall, R. et al.: A Student Response System for Increasing Engagement, Motivation, and Learning in High Enrollment Lectures. *AMCIS 2005 Proceedings*, S. 621–626, 2005.
- [Ha14] Hallmann, C.: Spielst du noch oder lernst du schon? - Spielend lernen in der Vorlesung. http://www.hochschullehre.org/wp-content/files/ZHW-Almanach-Paderborner-Beitrag-2014-03-Hallmann-Spielend_lernen.pdf, Stand: 23.02.2017.
- [Ho15] Hobert, S. et al.: Supporting Learner-Content Interaction on Autodidactic Field Trips through Mobile Learning Applications. *AMCIS 2015 Proceedings*, S. 1–12, 2015.
- [LSS14] Lucius, K.; Spannagel, J.; Spannagel, C.: Hörsaalspiele im Flipped Classroom. In (Rummler, K., Hrsg.): *Lernräume gestalten - Bildungskontexte vielfältig denken*. Waxmann, Münster [u.a.], S. 363–376, 2014.
- [Lu14] Lucius, K.: #20 Flipped twenty-five, <https://kristinalucius2.wordpress.com/2014/04/02/20-flipped-twenty-five/>, Stand: 23.02.2017.
- [Sa13] Sailer, M. et al.: Psychological Perspectives on Motivation through Gamification. *Interaction Design and Architecture(s) Journal* 19, S. 28–37, 2013.
- [SG05] Schwabe, G.; Göth, C.: Mobile learning with a mobile game. Design and motivational effects. *Journal of Computer Assisted Learning* 3, S. 204–216, 2005.

Eine Middleware Infrastruktur für das Teilen von Lerndaten und Diensten zwischen Bildungseinrichtungen

Francois Dubois¹, Truong-Sinh An¹, Agathe Merceron¹

Abstract: Im Projekt “Smart Learning im Handwerk” (SLHw) ist vorgesehen, dass mehrere Bildungseinrichtungen den gleichen Kurs anbieten und dabei dieselben Lernmaterialien nutzen können. Ferner sollen Benutzerinteraktionen mit dem Material lokal gespeichert werden. Um diesen Anforderungen gerecht zu werden, wurde eine Middleware Infrastruktur entwickelt, welche Schnittstellen bereitstellt, um die Kommunikation zwischen Bildungsanbietern und die Nutzung von Diensten zu erlauben, ohne dass diese Bildungsanbieter oder die Dienste direkt miteinander kommunizieren. Dieser Beitrag präsentiert diese Middleware Infrastruktur, ihren Einsatz mit drei Bildungseinrichtungen und vier Kursen sowie eine erste technische Evaluation.

Keywords: Middleware Infrastruktur, serviceorientierte Architektur (SOA), Application Programming Interface (API), Datenmodell, Virtuelle Organisation.

1 Einleitung

Digitale Lernplattformen sind essentiell für Bildungseinrichtungen; insbesondere beim lebenslangen Lernen sind diese ein wesentlicher Bestandteil und gewinnen zunehmend an Bedeutung. Eine Bildungseinrichtung kann beispielsweise eine Universität, eine Schule oder ein externer Dienstleister sein, der Bildungsangebote bereitstellt. In diesem Beitrag wird eine Handwerkskammer (HwK) als externer Dienstleister aufgeführt. Während die ersten digitalen Lernplattformen eine monolithische Architektur zugrunde legten, werden vermehrt auch modulare und serviceorientierte Plattformen entwickelt. Die große Nutzer-Community der weit verbreiteten Moodle Lernplattform stellt immer neue Module bereit, welche in Moodle integriert werden können. Ein Modul, das für Moodle entwickelt wird, kann nicht ohne Weiteres in eine andere Lernplattform integriert werden. Im Zuge der Weiterentwicklung von Moodle und dem vermehrten Einsatz mobiler Endgeräte wird neben dem modularen Aufbau die Integration von serviceorientierten Ansätzen hin zu einer “next generation platform” [Da07] verfolgt.

Es ist nicht selten, dass eine Bildungseinrichtung mehrere Lernplattformen nutzt, oder dass Lehrende an mehreren Bildungseinrichtungen lehren, die verschiedene Lernplattformen einsetzen, wie es insbesondere im Rahmen des lebenslangen Lernens und bei Kooperationen unter diesen auftritt. Um die Kommunikation zwischen den Lernplattformen zu fördern - z. B. der Austausch von Lernmaterialien - wurden Spezifikationen wie LTI entwickelt. Jede Lernplattform, die LTI-Consumer-fähig ist,

¹ Beuth Hochschule für Technik Berlin, Luxemburger Straße 10, 13353 Berlin,
{vorname}.{nachname}@beuth-hochschule.de

kann externes Lernmaterial präsentieren, welches der LTI-Spezifikation folgt. Im Projekt “Smart Learning im Handwerk - SLHw” [Kr16a] wird der Ansatz verfolgt, dass Lernmaterialien, die für eine Handwerkskammer (HwK) entwickelt wurden, auch von anderen HwK genutzt werden können, mit der Besonderheit, dass die Partner weiterhin ihre eigenen Lernplattformen einsetzen. Hierbei müssen es nicht die gleichen Lernplattformen sein. Ferner sieht das Projekt die Entwicklung einer *Recommendation Engine* [Kr16b] und einer *Learning Analytics Platform* [An16] vor. Diese Dienste sollen prinzipiell auch für alle HwK zur Verfügung stehen. Zur Realisierung müssen Benutzerinteraktionen zu dem Lernmaterial erfasst werden. Die Speicherung muss wiederum lokal unter Hoheit der jeweiligen HwK erfolgen, da diese über eine eigene Nutzerverwaltung verfügt. Die Hoheit aus Sicht eines Nutzers, wird dann wirksam, wenn Daten mehrerer Bildungseinrichtungen durch eine Anwendung zentralisiert werden. In diesem Fall muss der Nutzer für jede Bildungseinrichtung dem Anfordern von Daten zustimmen, indem Zugangsdaten bereitgestellt werden.

Diese Anforderungen entsprechen nicht der Funktionalität gängiger Lernplattformen. Um diesen Anforderungen gerecht zu werden, wurde eine Middleware Infrastruktur (MI) entwickelt, welche Schnittstellen bereitstellt, um die Kommunikation zwischen Bildungsanbietern und die Nutzung von Diensten zu erlauben, ohne dass die Bildungsanbieter oder die Dienste direkt miteinander kommunizieren. Somit ist das Integrieren einer weiteren Bildungseinrichtung oder eines neuen Dienstes mit minimalem Aufwand möglich. In diesem Beitrag wird die MI vorgestellt. Dabei wurde Wert auf die Verwendung üblicher Standards und Spezifikationen im E-Learning gelegt. Dies führte zur Entwicklung mehrerer Editoren, um insbesondere Übungen nach der QTI-Spezifikation und Lernmaterial nach der IMS-Spezifikation zu erstellen, da es keine quelloffenen Editoren am Anfang des Vorhabens gab, die Standards und Spezifikationen umsetzen, wie sie im Projekt erforderlich sind.

Der Beitrag ist wie folgt organisiert. Der nächste Abschnitt stellt vorangegangene Arbeiten dar. Im Abschnitt 3 werden die Middleware Infrastruktur (MI) sowie das Repository mit den Editoren eingeführt. Abschnitt 4 präsentiert den aktuellen Einsatz der MI als auch eine technische Evaluierung. Der Beitrag endet mit einer Zusammenfassung und Diskussion.

2 Vorangegangene Arbeiten

Bei der Entwicklung einiger Lernplattformen werden Anwendungsdienste ähnlich des IMS Abstract Framework [IM03], in eigenständige Komponenten aufgeteilt. Mit diesem Ansatz ist es beispielsweise möglich, die Kurs- und Nutzerverwaltung von einem Repository für Lerninhalte zu kapseln und in eine Lernplattform zu integrieren. Dabei kann es sich um eine serviceorientierte Architektur (SOA) handeln. Hierbei wird die infrastrukturelle Kommunikation zwischen verschiedenen Anbietern (Provider) und Konsumenten (Consumer) von Dienstleistungen (Services) geregelt.

Klassisch wird in diesem Zusammenhang oftmals das find-bind-execute- Paradigma angewendet [Zh05]. Hierbei stellt ein Service Consumer eine Anfrage an einen Service Broker. Dieser hat als Vermittler die Aufgabe, einen passenden Service Provider aufzufinden (find). Servicebeschreibungen registrierter Service Provider sind zu diesem Zweck zentral in einer Service Registry hinterlegt. Sobald ein Service Consumer eine passende Servicebeschreibung von dem Service Broker erhalten hat, kann dieser eine Verbindung zu einem Service Provider herstellen (bind). Abschließend erfolgt das Ausführen einer Dienstleistung (execute).

Das folgende Beispiel illustriert, wie Service Consumer und Service Provider in der E-Learning-Branche aussehen können. Im Projekt SLHW wurde eine *Learning Companion App* (LCA) [Kr16a] entwickelt, die Lerninhalte und Empfehlungen für Lerninhalte präsentiert. Diese App benötigt den Zugang zu einem Repository, von dem die Lerninhalte angefordert werden. Die LCA kann als Service Consumer aufgefasst werden. Diese nutzt die Dienstleistungen des Repositories, welche von einer Bildungseinrichtung als Service Provider bereitgestellt werden.

Die Problematik, unter welchen Bedingungen und Restriktionen ein Service Provider Dienstleistungen teilen kann, wird unter anderem im Bereich des Grid Computing diskutiert. Ein Zusammenschluss von Service Providern (Bildungseinrichtungen) für eine zentrale Bündelung von Ressourcen (Lerninhalten) wird in diesem Zusammenhang als Virtuelle Organisation bezeichnet [FK03]. In der E-Learning-Branche wurde das Konzept des Grid-Computing bereits in der Architektur E-Learning Grid erprobt [IJ14]. Durch die Zentralisierung von Service Providern könnten zudem interuniversitäre Anwendungsszenarien realisiert werden, wie beispielsweise gemeinsame Rahmenpläne und Lerninhalte, auch bekannt als Joint-/Double-Degree [AQS08]. Die hier vorgestellte Middleware Infrastruktur basiert auf dem SOA-Paradigma [Zh05] und Microservices [FL15]. Diese Ansätze erlauben ähnliche Anwendungsszenarien wie im E-Learning Grid [IJ14, S. 131].

In [Le08] werden verschiedene Formen des rechnergestützten Lernens vereint. Exemplarisch verknüpft eine serviceorientierte Architektur Multimediageräte für die Video- und Audioaufnahme einer Präsenzveranstaltung sowie die virtuelle Welt “Second Life” als auch das Lernmanagementsystem Stud.IP. Die Grundidee hierbei ist eine Präsenzveranstaltung, um eine virtuelle Veranstaltung zu erweitern. Die dazu benötigten Webservices werden klassisch, wie auch in diesem Beitrag über einen Service Broker, verwaltet. Besitzt eine Anwendung keine Webservice-Schnittstellen, wie beispielsweise Second Life, werden diese über sogenannte Surrogates adaptiert und repräsentiert. Das Adaptieren von Protokollen und Transferdaten wird im Kapitel Infrastruktur näher erläutert.

Eine ähnliche SOA wird in [ZT09] vorgestellt. Besonders hervorzuheben ist hierbei die Herangehensweise Kontext-differenzierte Schnittstellen einer Universität, wie die Administration (Personal, Räume, Termine), die Wissensverwaltung (Archive) und die Lernsysteme zu betrachten. Ein Vorteil von SOA besteht darin, heterogene Systeme zu

orchestrieren. Auf diese Weise können die bereitgestellten Dienstleistungen für Arbeits-, Informations- und Lernprozesse zentralisiert werden. Der Hauptunterschied zwischen dem Ansatz in [ZT09] und dem Ansatz in diesem Beitrag liegt in der Beschreibung der Dienste, welche die Infrastruktur bereitstellt. In [ZT09] wird speziell ein XML-basiertes Format entwickelt: Service Technology-independent Language (STiL). Jeder Dienst wird in STiL repräsentiert. In diesem Beitrag werden Dienste als Microservices beschrieben. Jeder Microservice verfügt über eine REST-API. Sofern vorhanden, verwenden die Schnittstellen und Datenmodelle Standards oder verbreitete Spezifikationen, wie LTI für Lehrmaterialien oder xAPI für Benutzerinteraktionen.

Ein weiterer Ansatz nach [MH12] ist die Konzeption von Lernmanagementsystemen durch die Verwendung von Cloud-Lösungen. Besonders bei Bildungseinrichtungen, die keine Infrastruktur oder Lerninhalte bereitstellen können ist dies attraktiv. Das Teilen oder die Nutzung von Lerninhalten erfolgt über die Cloud. Im Vergleich zu diesem Beitrag ist die Datenhoheit, trotz möglicher Zugriffseinschränkungen einer Bildungseinrichtung, nicht im Fokus, da die Daten extern verwaltet werden. Die Middleware Infrastruktur verwaltet keine Daten und regelt nur den Zugriff. Zudem wird bei einer Cloud-Lösung weniger die Heterogenität der Systeme betrachtet. Vielmehr verfolgt die Middleware Infrastruktur den Ansatz, etablierte Systeme innerhalb einer Bildungseinrichtung zu integrieren, was eine Kooperation oder das Teilen von Lerninhalten durch beliebiges Orchestrieren der Dienstleistungen nicht ausschließt.

3 Infrastruktur

Die Grundidee der Middleware Infrastruktur (MI) und deren wichtigsten Komponenten werden im Folgenden exemplarisch erläutert. Zwei Bildungseinrichtungen B1 und B2 nutzen die eigenen Lernplattformen B1_L1 und B2_L1 für die Verwaltung ihrer Nutzer. Zudem sollen Kurse aus deren lokalen Repositories (B1_R1, B2_R1) den Lernenden zur Verfügung gestellt werden. Die Lernenden verwenden die *Learning Companion App*, welche die Kurse einer Lernplattform bereitstellt. In diesem Szenario sind beide Lernplattformen und Repositories als Service Provider zu betrachten und die *Learning Companion App* als Service Consumer. Der Service Broker mit Hilfe der Service Registry ermöglicht es zu entscheiden, von welcher Bildungseinrichtung, B1 oder B2, die Kursmaterialien angefordert werden. Da die Kursmaterialien aus verschiedenen Repositories angefordert werden, können potenziell unterschiedliche Datenmodelle vorliegen. Zu diesem Zweck hat die MI die Aufgabe, ein einheitliches Datenmodell bereitzustellen, in dem die Datenmodelle der Service Provider transformiert werden. Auf diese Weise kann gewährleistet werden, dass eine Interoperabilität für Service Consumer entsteht. Nachfolgend werden die Begrifflichkeiten von SOA auf die Komponenten der MI überführt.

3.1 Middleware und Management

Aus technischer Sicht lässt sich die Middleware Infrastruktur in die fünf Systemkomponenten Third Party Application (TPA), Third Party Module (TPM), Middleware API (MWA), Hub und Client Application (CA) unterteilen.

Eine TPA ist vergleichbar mit einem Service Provider. Hierbei könnte es sich um Repositories mit Lerninhalten, Learning Record Stores zum Speichern von Benutzerinteraktionen oder Learning Management Systemen von Bildungseinrichtungen handeln.

Ein Third Party Module ist ein Adapter, der ein Datenmodell einer TPA in einen einheitlichen Standard transformiert, um eine interoperable Distribution zu ermöglichen. Als Beispiel transformiert das TPM eine Kursstruktur aus der TPA B1_R1 in ein Datenmodell, das an die Spezifikation IMS CC angelehnt ist (siehe Abbildung 1). Als weiteres Beispiel werden die Benutzerinteraktionen mit dem Lernmaterial als Statements nach der xAPI-Spezifikation transformiert. In Abbildung 1 sind exemplarisch lokale Learning Record Stores (RS) B1_RS1 und B2_RS1 der jeweiligen Bildungseinrichtungen B1 und B2 zum Speichern dieser xAPI-Statements aufgeführt.

Eine Middleware API (MWA) stellt einheitliche Vermittler-Schnittstellen für Dienstleistungen verschiedener TPA bereit und kann mit einem Service Broker verglichen werden, mit der Besonderheit, dass die Schnittstellen durch die Datenmodelle der Third Party Module definiert sind. An dieser Stelle wird das Konzept der Microservices [FL15] aufgegriffen, da flexible Anforderungen in der E-Learning-Branche, wie beispielsweise persönliche Lernumgebungen, zu berücksichtigen sind. Das Konzept der Microservices wird dadurch realisiert, dass es nicht eine einzige API (Application Programming Interface) gibt, sondern mehrere. Aktuell unterteilen sich die MWA nach Kurs- und Nutzerverwaltung (LMS), Aufzeichnen von Lernaktivitäten (LRS), Verwalten von Lerninhalten (LCMS) und Kommunikation zwischen Lernenden und Lehrenden durch Foren (COM, in Abbildung 1 nicht aufgeführt). Die COM-MWA versucht, hierarchische Kommunikationsstrukturen abstrakt abbilden zu können, was aufgrund einer fehlenden Standardisierung die grundlegende Adaptierung für verschiedene Foren ermöglichen könnte. Die verwendeten Entitäten sind Forum, Diskussion und Beitrag. Ein Beitrag kann durch Antworten beliebig verschachtelt werden. Ob eine Diskussion initial einen Beitrag voraussetzt, obliegt der Drittanbieteranwendung.

Jede MWA agiert unabhängig, wodurch auch neue Dienstleistungen nachträglich durch die Einführung weiterer MWA integriert werden können.

Ein pervasiver Ansatz, wie in [ZT09] erfordert, dass jede MWA unabhängig von der Drittanbieteranwendung (TPA) keine erneute Authentifizierung durch den Lernenden benötigt. Bei diesem Middleware Ansatz wird vorausgesetzt, dass die TPA ein Authentifizierungsverfahren wählen, mit dem es ermöglicht wird den Authentifizierungsprozess, wie z. B. bei Single-Sign-On (SSO), zu zentralisieren.

Eine Client-Anwendung (CA) beansprucht die Dienstleistungen einer MWA, um mit einer TPA verbunden zu werden. Im Rahmen des SLHw-Projekts werden insbesondere die *Learning Companion App* und die Editoren als CA integriert (siehe Abbildung 1). Welche TPA von der CA angefordert wird, ist durch sogenannte Routen gewährleistet, siehe Systemkomponente Hub im nächsten Absatz. Um Anforderungen des lebenslangen Lernens und der persönlichen Fortbildung durch Personal Learning Environments zu berücksichtigen, sind CA prinzipiell in der Lage, Dienstleistungen verschiedener MWA beliebig zu verknüpfen (Orchestration), um neue Lernerfahrungen bieten zu können. Es ist hervorzuheben, dass die MI kein Datenaustausch zwischen den TPA initiiert, was die Datenhoheit und Datentransparenz für alle Beteiligten begünstigt. Daraus folgt, dass datenschutzrechtliche Aspekte der E-Learning-Branche durch das verteilte System nicht ausgeweitet werden.

Der Hub ist für das Verwalten von CA-Eigentümern (CA Owner in Abbildung 1) sowie Routen verantwortlich und erfüllt die Aufgaben einer Service Registry. In jeder Route wird hinterlegt, welche Zugangsdaten für eine Third Party Application benötigt werden und wie die TPA zu erreichen ist. Für jeden CA-Eigentümer wird überprüft, welche Routen verwendet werden dürfen.

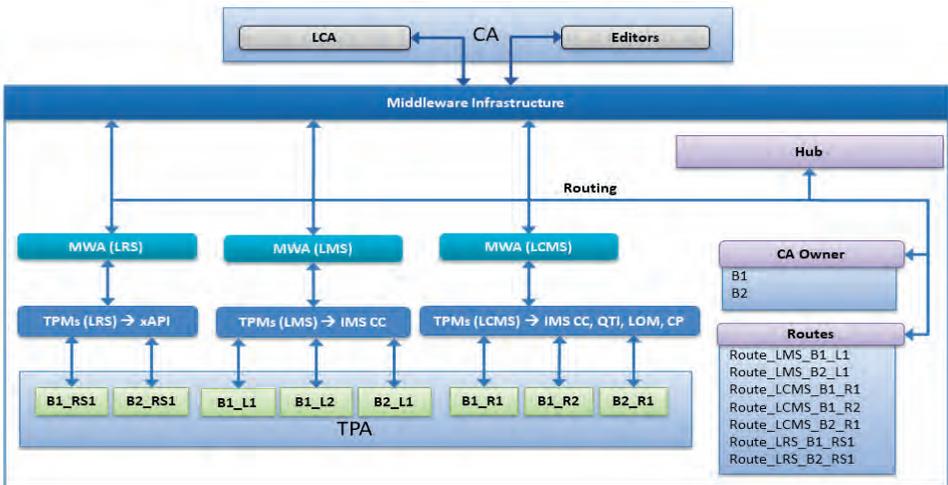


Abb. 1: Middleware Infrastruktur

3.2 Technische Umsetzung

Die Middleware Infrastruktur wurde mit der Softwarearchitektur Java EE entwickelt. Integriert wurde vor allem das Java Naming and Directory Interface (JNDI). Mithilfe dieser Programmierschnittstelle können Objektreferenzen unter Berücksichtigung der Objekt-ID in einer Registry für die spätere Nutzung abgelegt werden. Bei der Middleware Infrastruktur handelt es sich bei den Objektreferenzen um Data Access

Objects (DAO) eines Third Party Modules. Die ID zum Auffinden eines DAO wird dynamisch durch die Attribute einer Route ermittelt. Eine Route ist lediglich ein fest definiertes Datenmodell nach der JavaScript Object Notation (JSON), das den technischen Verbindungsweg zu einer Bildungseinrichtung beschreibt. Zudem werden die für den Verbindungsaufbau benötigten Parameter hinterlegt, wie beispielsweise die Zugangsdaten. Eine Middleware API (MWA) ist nicht auf ein Netzwerkprotokoll festgelegt. Die Schnittstellen für den JNDI-Lookup können beliebig innerhalb eines Services eingesetzt werden. Es empfiehlt sich jedoch zugunsten der Nutzbarkeit und Standardisierung alle Schnittstellen einer MWA zu vereinheitlichen. Aktuell werden alle MWA basierend auf der JAX-RS-Spezifikation (REST) durch RestEasy umgesetzt. Das Deployment erfolgt über einen Wildfly-Server.

Die Middleware Infrastruktur ist Modular konzipiert. Dies bedeutet, dass Hub, TPA und TPM Bestandteil der Middleware sind. Das Ziel ist es, die nachträgliche Integration von Systemen und Dienstleistungen durch eine Community zu gewährleisten. Einer der größten Vorteile einer SOA ist die stetige Erweiterbarkeit. In diesem Zusammenhang besitzt die Middleware eine logische Codestruktur, die eine klare Trennung von Entwickleraufgaben sicherstellt. Es ist zu vermeiden, dass ein Entwickler Programmcode verstehen oder bearbeiten muss, der für die Lösung des Problems irrelevant ist. Es werden die Kernfunktionalitäten in Form von funktionalen Modulen vorgestellt. Ein funktionales Modul ist ein Versuch, auf Codeebene eine lose Kopplung von Funktionalitäten zu realisieren. Fast alle funktionalen Module definieren Schnittstellen oder Annotationen für Entwickler, um die Komplexität der Middleware Infrastruktur zu verbergen. Hierbei wird das Umsetzungsparadigma Inversion of Control (IoC) nach [Fo04] berücksichtigt. Die Abbildung 2 bietet eine Übersicht der funktionalen Module. Zudem wird die Integration domänenspezifischer Lösungen mit einbezogen.

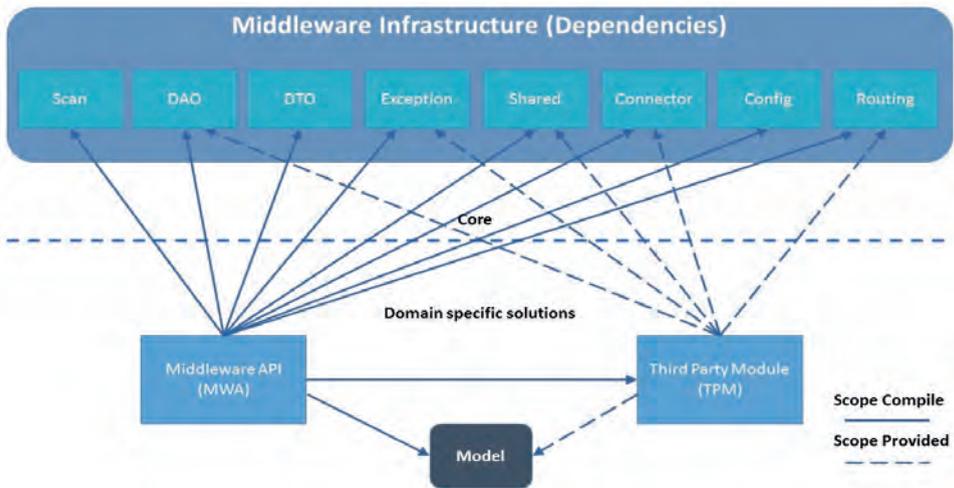


Abb. 2: Middleware Infrastruktur (Abhängigkeiten)

Tab. 1: Funktionale Module

Beschreibung	Module
Auflösen von verfügbaren TPM in einer MWA	Scan
Aufsuchen der Data Access Objects 1	DAO
Transferobjekte für einen einheitlichen Datenaustausch	DTO
Fehlerbehandlung der Middleware Infrastruktur	Exception
Util-Klassen und geteilte Datenmodelle	Shared
Verbindungstypen der Third Party Modules	Connector
Logik zum Einlesen einer Konfigurationsdatei	Config
Anfragen von Routen	Routing

Innerhalb eines TPM werden alle funktionalen Module aus der Tabelle 1 über das Build-Management-Tool Maven mit dem Scope `provided` versehen. Das bedeutet, dass nur die MWA, die im späteren Verlauf beliebig viele TPM einbindet, Abhängigkeiten zur Verfügung stellt. Ein TPM nutzt die funktionalen Module der MWA.

3.3 Repository und Editoren.

Im Projekt SLHw wurde für die Ablage von Lernobjekten (LO) und Kursstrukturen ein Repository als TPA geschaffen. Zudem wurden zwei CA zum Verwalten und Präsentieren von LO entwickelt (vgl. Editors in Abbildung 1). Das Autorentool `openStudio` [An16] ist eine Webapplikation und ermöglicht durch diverse Editoren das Erstellen und Pflegen von LO – inklusive QTI-Aufgaben – und deren Metadaten sowie das Komponieren von LO in Lerneinheiten. Das Präsentationstool `openPresentation` [An16] visualisiert unter anderem die LO, welche durch die Dienstleistungen der LCMS-Middleware API von einem Repository (siehe B1_R1 oder B2_R1 in Abbildung 1) angefordert wurden. Die LO in dem Projekt SLHw sind die kleinste wiederverwendbare Einheit mit eigenen Metadaten nach der Spezifikation IMS LOM. Projektbezogen benötigt jedes LO Metadaten mit mindestens einem Lernziel und kann jede Form von Rich Media Typen annehmen. Ein LO kann demnach ein Video, eine Animation, ein Dokument oder eine Übung nach der Spezifikation IMS QTI sein. Technisch gesehen ist ein LO jedoch eine Kapselung verschiedener Typen nach der Spezifikation IMS CP.

4 Einsatz und Evaluierung

Das Szenario mit mehreren Handwerkskammern, welche auf dieselben Lernobjekte (LO) aus einem einzigen Repository zugreifen, wurde getestet, aber mit realen Lernenden aus organisatorischen Gründen nicht umgesetzt. Es wurde das Szenario von drei Bildungseinrichtungen realisiert, welche gemeinsam die *Learning Companion App* als Client-Anwendung benutzen, um Zugang zu vier Kursen zu ermöglichen.

4.1 Anwendung: 4 Kurse von 3 Bildungseinrichtungen

Im Rahmen des Projektvorhabens wurde der Fortbildungslehrgang “Gebäudeenergieberater” an der Handwerkskammer Berlin im Herbst 2016 mit einer Gesamtdauer von 24 Wochen durchgeführt. Weiterhin wurde die gesamte Infrastruktur auch im Wintersemester 2016/2017 an der Technischen Universität Berlin (TU-Berlin) und an der Beuth Hochschule für Technik Berlin (Beuth-HS) erprobt. An dem Kurs “Advanced Web Technologies” im Studiengang Master Informatik der TU-Berlin haben sich 142 Studenten über 12 Wochen kursbegleitend auf die Prüfung vorbereitet. Im Kontrast dazu wurde der Kurs “JavaFX” der Beuth-HS allen Online-Studenten im Studiengang Bachelor Medieninformatik über einen Zeitraum von 10 Wochen zur Verfügung gestellt. Die Teilnahme an dem Kurs war freiwillig und enthielt keine prüfungsrelevanten Leistungen. Es beantragten 53 Interessierte einen Zugang zu dem Kurs. Weiterhin wurde ein “Fitnesskurs” von der Handwerkskammer Berlin zu Demozwecken angeboten.

Drei der bisherigen Kurse starteten zu unterschiedlichen Terminen und liefen über eine Phase von 3 Monaten parallel. Hierbei betrieb jede Bildungseinrichtung ein eigenes Repository. Die insgesamt über 2400 LO wurden von den Bildungseinrichtungen mit openStudio eigenverantwortlich in die jeweiligen Repositories eingepflegt. Des Weiteren verwaltet jede Bildungseinrichtung mit deren Lernplattformen die eigenen Benutzer. Alle Kurse wurden den Benutzern über die *Learning Companion App* (LCA) bereitgestellt. Da die LCA einen Benutzer nicht einer Lernplattform zuordnen kann, spielt die MI mit ihren *Route*-fähigen Middleware APIs eine wesentliche Rolle bei der Kommunikation mit den entsprechenden TPA. Abstrakt beschrieben kommuniziert die LCA für die Benutzer- und Kursverwaltung mit der Lernplattform B1_L1 über die *Route* *Route_LMS_B1_L1*; für die Lerninhalte integriert die LCA das LTI-Tool openPresentation. Dieses kommuniziert mit dem Repository B1_R1 über die *Route* *Route_LCMS_B1_R1* (vgl. Abbildung 3, oben). Die Nutzer können auch die Lernplattform direkt als Zugangspunkt nutzen. In diesem speziellen Szenario integriert die Lernplattform B1_R1 das LTI-Tool openPresentation (vgl. Abbildung 3, unten).

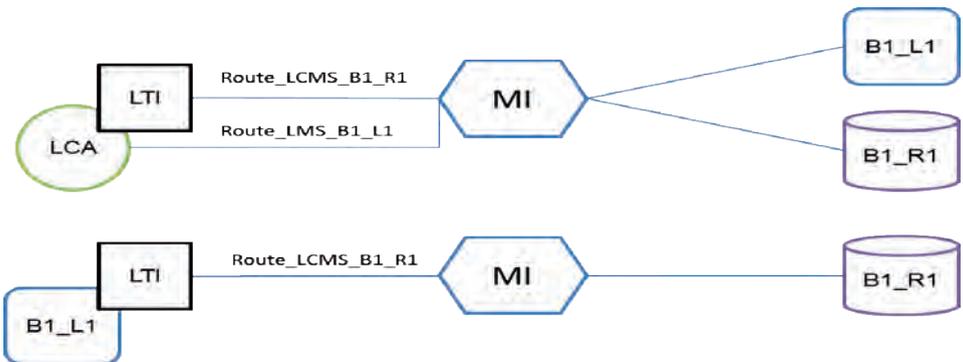


Abb. 3: Bildungseinrichtungen mit getrennten Repositories

Im Rahmen der Öffentlichkeitsarbeit wurde zusätzlich ein Demo-Setup (Lernplattform B1_L2, siehe Abbildung 1) aufgesetzt, bei dem Interessierten die Möglichkeit geboten wird, einen Ausschnitt des Fortbildungslehrgangs “Gebäudeenergieberater” oder einen “Fitnesskurs” zu testen. Hierbei liegt der Kurs “Gebäudeenergieberater” im Repository B1_R1 und der “Fitnesskurs” im Repository B1_R2. Die Interessierten können über die Demo Lernplattform B1_L2 den Kurs “Gebäudeenergieberater” und den “Fitnesskurs” aus den beiden Repositories B1_R1 und B1_R2 über die *Routen* Route_LCMS_B1_R1 und Route_LCMS_B1_R2 besuchen. Somit ist es möglich, dass über die Lernplattformen (B1_L1, B1_L2) auf denselben Kurs des Repository (B1_R1) zugegriffen werden kann. Ebenso ist es möglich, dass eine Lernplattform B1_L2 verschiedene Kurse aus unterschiedlichen Repositories (B1_R1, B1_R2) anfordert (siehe Abbildung 4).

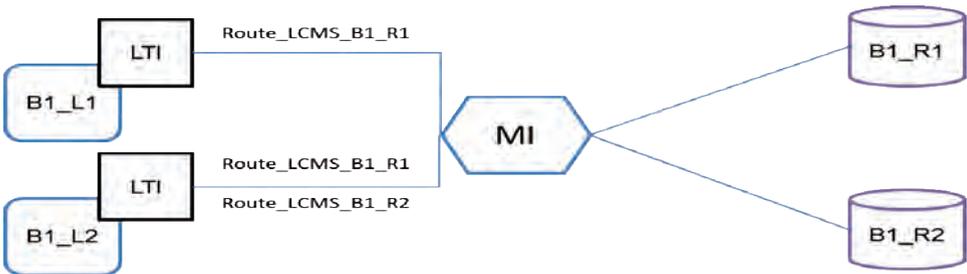


Abb. 4: Bildungseinrichtungen nutzen verschiedene Repositories

4.2 Technische Evaluation

Neben den realen Kursen erfolgte in einer kontrollierten Testumgebung eine vorläufige quantitative Evaluation der MI. Die Auswertung ist als Richtwert zu verstehen, da für Stresstests keine Produktionsumgebung gewährleistet werden konnte. Die Daten werden zudem durch geringe Latenzzeiten des Intranets verzerrt. Für alle Systemkomponenten wurden die gleichen technischen Spezifikationen gewählt. Jedes Testszenario erfolgte zeitlich versetzt 8-mal, um mögliche Abweichungen durch Störfaktoren zu minimieren. Das erste Ergebnis zeigt die Verarbeitungszeit der MWA, ohne Dienstleistungen einer TPA anzufordern. Bei 250 simultanen Clients mit jeweils 100 sequenziellen Anfragen beträgt die Laufzeit eines Requests im Durchschnitt 0,48 Sekunden. Das zweite Ergebnis ermittelte den Datendurchsatz der MI im Vergleich zu einem direkten Aufruf einer TPA. Der Overhead der MI hinsichtlich des Datendurchsatzes ergibt bei 50 simultanen Clients mit jeweils 250 sequenziellen Anfragen in der kontrollierten Testumgebung bei Moodle (TPA) eine Differenz von etwa 3%. Als Hauptfaktoren, die maßgeblich die Performance beeinflussen, konnte die Netzwerkinfrastruktur sowie die Performance der einzelnen TPA ermittelt werden. Zweitrangig ist die Datentransformation und das Auffinden eines TPM. Generiert und erfasst wurden die Testdaten mit Hilfe von Apache JMeter als CA Simulation und VisualVM für das Remote-Monitoring des Anwendungsservers der MI.

Für technisch versierte Leser ist anzumerken, dass ein Bottleneck durch die Verwendung eines Hubs ausgeschlossen wird, indem die Routen lokal in den MWA zwischengespeichert werden.

5 Zusammenfassung und Diskussion

In diesem Beitrag wurde die Middleware Infrastruktur vorgestellt, welche im Projekt SLHw entwickelt wurde. Diese serviceorientierte Infrastruktur erlaubt eine flexible Kommunikation und Mutualisierung von Diensten zwischen Bildungsanbietern. Sie wurde mit drei Bildungseinrichtungen und vier Kursen erfolgreich eingesetzt. Obwohl nur Szenarios im Bereich E-Learning vorgestellt wurden, ist die Middleware Infrastruktur unabhängig von der Anwendungsdomäne. Diese könnte beispielsweise mit Diensten und Third Party Applications aus dem Bereich E-Health erweitert werden: Angebot eines Fitness-Kurses mit Integration von Fitness-Armbändern. Voraussetzung ist, dass passende Third Party Modules und Middleware APIs entwickelt werden.

Die Middleware Infrastruktur hat Potenzial für virtuelle Organisationen, wie z.B. die Virtuelle Fachhochschule (VFH) [Vi17]. In der VFH bieten mehrere Hochschulen gemeinsam Online-Studiengänge an. Jede Hochschule ist für die Entwicklung von digitalen Medien in einigen Kursen verantwortlich. Mit der Autorensoftware *LOOP* (Learning Object Online Plattform) können die digitalen Medien entweder direkt in der zentralen Ablage oder an jedem Standort lokal entwickelt und gepflegt werden. Anschließend werden diese in das zentrale Repository übertragen. Die MI würde ermöglichen, dass die digitalen Medien aus den lokalen Repositories der jeweiligen verantwortlichen Hochschule zu einem einzigen Kurs virtuell zusammengeführt werden.

Es ist geplant, die Middleware Infrastruktur, das Repository, die Editoren und LTI-Tools zum Ende des Projektes im September 2017 quelloffen zu veröffentlichen.

Danksagung: Wir danken B. Bayat, M. Bärenfänger, R. Chandru, M. Dinziol, M. Edriss, I. Fritsch, H. John, D. Jürgensen, J. Kania, C. Krauß, E. Manthey, U. Meurer, S. Müller, M. Scharp und M. Zwicklbauer für Diskussionen. Das diesem Beitrag zugrundeliegende Vorhaben wird mit Mitteln des BMBF Förderkennzeichen 01PD14002B gefördert. Die Verantwortung für den Inhalt liegt bei den Autoren.

Literaturverzeichnis

- [An16] An, T.-S.; Dubois, F.; Manthey, E.; Merceron, A.: Digitale Infrastruktur und Learning Analytics in Co-Design. In: Proceedings of the Workshop Learning Analytics, co-located with the 13th e-Learning Conference of the German Society for Computer Science, Potsdam, Germany, September 11, 2016, <http://ceur-ws.org/Vol-1669/>, 2016.

- [AQS08] Aguirre, S.; Quemada, J.; Salvachúa, J.: Work in progress - Developing Joint Degrees through e-Learning systems. Proceedings - Frontiers in Education Conference, FIE, ISSN 15394565. doi: 10.1109/FIE.2008.4720474, S. 19–20, 2008.
- [Da07] Dagger, D.; O'Connor, A.; Lawless, S.; Walsh, E.; Wade, V. P.: Service-Oriented e-learning platforms: From monolithic systems to flexible services. IEEE Internet Computing, ISSN 10897801. doi: 10.1109/MIC.2007.70, S. 28–35, 2007.
- [FK03] Foster, I.; Kesselman, C.: The Grid 2: Blueprint for a New Computing Infrastructure. Elsevier Ltd., 2 edition, ISBN 978-1558609334, 2003.
- [FL15] Fowler, M.; Lewis, J.: Microservices: Nur ein weiteres Konzept in der Softwarearchitektur oder mehr? Objektspektrum, S. 1–7, 2015.
- [Fo04] Fowler, M. 2004. Inversion of Control Containers and the Dependency Injection pattern, <https://www.martinfowler.com/articles/injection.html#>, Stand: 13.02.2017.
- [IJ14] Ivanovic, M.; Jain, L. C.: E-Learning Paradigms and Applications: Agent-based Approach. Springer, ISBN 978-3-642-41964-5. doi: 10.1007/978-3-642-41965-2, 2014.
- [IM03] IMS Global Learning Consortium. IMS Abstract Framework: White Paper, 2003, <https://www.imsglobal.org/af/afv1p0/imsafwhitepaperv1p0.html>. Stand: 12.02.2017.
- [Kr16a] Krauss, C.; Merceron, A.; An, T.-S.; Zwicklbauer, M.; Arbanowski, S. The Smart Learning Approach - A mobile Learning Companion Application. In: Proceedings of The Eighth International Conference on Mobile, Hybrid, and On-line Learning (eLmL 2016). IARIA, April 24 - 28, Venice, Italy, 2016.
- [Kr16b] Krauss, C.: Smart Learning: Time-Dependent Context-Aware Learning Object Recommendations. Proceedings of the 29th International Florida AI Research Society Conference (FLAIRS-29), AAAI, Key Largo, S. 501-504, 2016.
- [Le08] Lehsten, P.; Thiele, A.; Zilz, R.; Dressler, E.; Zender, R.; Lucke, U.; Tavangarian, D.: Dienste-basierte Kopplung von virtueller und Präsenzlehre. In: DeLFI 2008, S. 77-88, 2008.
- [MH12] Masud, M. A. H.; Huang, X.: An e-learning system architecture based on cloud computing. system, 10/11, S. 736-740, 2012.
- [Vi17] Virtuelle Fachhochschule, <https://www.vfh.de>. Stand: 20.03.2017.
- [ZT09] Zender, R.; Tavangarian, D.: Service-oriented university: Infrastructure for the university of tomorrow. In: Intelligent Interactive Assistance and Mobile Multimedia Computing, S. 73-84. Springer Berlin Heidelberg, 2009.
- [Zh05] Zhu, H.: Challenges to reusable services. In: Proceedings - 2005 IEEE International Conference on Services Computing, SCC 2005, ISBN 0769524087. doi: 10.1109/SCC.2005.36, S. 243–244, 2005.

Moving freely while staying on track — Smart Glasses to Support Lecturers

Christian Wolters¹, Daniel Wessel¹, Finn Jacobsen¹ und Michael Herczeg¹

Abstract: This paper explores the use of smart glasses as a supporting tool for lecturers. Typical constraints of lecture rooms, like fixed lecture stands and projection surfaces, impose limits on a lecturer's interaction with the audience. Wearable devices like smart glasses could allow the lecturer to move freely, keep continuous eye-contact to facilitate attention, while providing unobtrusive access to time information and lecture notes. Following a human-centered design process (HCD) the development and evaluation of a working prototype is presented. Results of the HCD process show the potential and feasibility of the proposed design solution, but also highlight the limitations of the current technology.

Keywords: face-to-face teaching, lectures, wearables, smart glasses, Google Glass

1 Introduction

The physical room of a face-to-face lecture is still very important in education. However, in many lectures, the lecturer uses a data projector to visualize content. While convenient, the use of a computer on a fixed lecture stand requires the lecturer to stay in one place or frequently return to it, e.g., to advance the slides or look at lecture notes to stay on track (e.g., using the Presenter Display in PowerPoint or Keynote). Using the lectern “hides” most of the lecturer while looking down on the notes breaks eye contact with the audience. Although remotes allow advancing slides from any location, they do not provide lecture notes. Smartphone apps like Apple’s Keynote iOS App do provide this information, but require the lecturer to look down on the small screen and thereby also breaking contact with the audience.

Time keeping is another challenge — considering topics in a lecture can take longer than expected if the students have questions or topics can be skipped, if the students already know about it. While clocks are in plain sight in many lecture rooms, they do not provide direct feedback when, e.g., half of the time is up, or when there are only five minutes left. The lecturers have to remember to look at the clock, calculate the remaining time and match the remaining content with the remaining time. Again, a presentation view of Keynote or PowerPoint can provide this information, but again only at the lectern.

Given these constraints of the lecture room, we look at ways to improve its infrastructure

¹ Institut für Multimediale und Interaktive Systeme, Universität zu Lübeck, Ratzeburger Allee 160, D-23562 Lübeck, {wolters|wessel|jacobsen|herczeg}@imis.uni-luebeck.de

in order to increase presence and eye contact with the students. We want to enable the lecturer to move freely and keep in contact with the audience to facilitate attention, while having unobtrusive access to lecture notes to stay on track and on time.

For this purpose, wearable devices like smart glasses seem very well suited. They are carried by the lecturer who can move freely. Smart Glasses, like Google Glass [St13], could display information about the presentation in real-time within the field of view of the lecturer (cf. Fig 1) while maintaining eye-contact with the audience.

But is this really the case? To address this question, we ask:

1. Can lecturers be supported by smart glasses like Google Glass?
2. What are the strengths and weaknesses of smart glasses in a lecture setting?

The study follows an ISO 9241-210 human-centered design process (HCD, [Pr11]), which was used to develop a working prototype. Steps in the process include understanding and specifying the context of use, specifying the user requirements, developing a prototype, and conducting multiple formative evaluations and a main summative evaluation of the prototype.



Fig 1: Google Glass Explorer Edition

When dealing with the use of smart glasses in face-to-face higher education, it is important to keep in mind that Google Glass does not replace the infrastructure already in place. It augments it and provides the lecturer with more freedom and possibilities to interact more directly with the audience.

2 Technical Background and Related Works

Smart glasses are a subcategory of head-mounted displays (HMDs). HMDs have been a topic of research since the late 1960s [Su68]. In recent years, commercially available and less expensive devices have become available with different characteristics. Important differences between models are whether the screen is projected on one eye (e.g. Google

Glass) or both eyes (e.g. Epson Moverio series), and whether the image is projected in the line of sight (e.g. Epson Moverio series) or above the line of sight (e.g. Google Glass). For use in lecture settings — with focus on eye contact with the audience — Google Glass seems to be the best choice. Google Glass Explorer Edition [Go17a] (cf. Fig 1) projects the image on a prism in front of the right eye (with a display resolution of 640 x 360 pixels) and seems relatively unobtrusive. Its unobtrusiveness should make it well suited for keeping eye contact. Even when looking at the projected image, the lecturer is oriented towards the audience and does not need to return to the lectern.

Studies with smart glasses have already been conducted in different application domains, for example regarding use cases and usability challenges in the healthcare domain [MWH15] or exploring the use of smart glasses in the museum [Ma16]. These two settings exemplify main potentials of smart glasses: interacting freely with a mobile computer, e.g. hands-free in sterile environments, and displaying information in mobile contexts, including by means of augmented reality (AR).

The mentioned potentials are also highly relevant in the education domain, for example regarding the potential usage of Google Glass for training future scientists in laboratory work [Hu15]. The authors' findings show the potential of smart glasses when it comes to situated access to information — hands-free if needed — but also point out technical and usability limitations. One important limitation shown by previous studies is user input via speech recognition (e.g., [We14]). In an educational setting, different students can be speaking at the same time, which results in errors in speech recognition. The authors propose the use of head gestures as an alternative.

A study in the context of higher education by [EME16] also explored the use of Google Glass in a face-to-face lecture setting, with an application displaying live results from an audience response system. The study addresses some issues with Google Glass including lecturers wearing eyeglasses, interfering light sources in the environment, limits to the amount of information presented, difficulties with input controls, the necessity of habituation to the device, excessive heat generation, short battery life, and unstable internet connection. Conscious of these limitations, the authors nevertheless concluded for audience response systems that Glass has benefits in face-to-face lectures and increases interaction with the audience.

Looking at the prior work, smart glasses like Google Glass do seem to have potential in educational settings. However, they also come with limitations. The questions remain: Can smart glasses support lecturers? What are their strengths and weaknesses of smart glasses in this context?

3 Analysis

As first step in the HCD process, interviews with the target audience were conducted to assess design features and possible concerns, followed by a context analysis.

3.1 Interviews

To understand and specify the context of use, six interviews were conducted with lecturers and research assistants at a university. All interviewees had prior experience with Google Glass, ranging from short tryouts to thorough testing and app development. Questions assessed lecturing habits, Google Glass, and ideas to support lectures with Google Glass.

Results regarding lecturing habits showed that the interviewees used MS PowerPoint or Apple Keynote, some in combination with remotes. No one used additional smartphone applications. Lecture notes were usually used for scientific presentation, not for lectures. However, they were open to trying out lecture notes if they would work with Google Glass.

The hardware itself was seen critically — with low display resolution, battery life and heat generation being the main issues. However, its potential was regarded as high enough to be willing to try a lecture app on Google Glass, although not yet in real lectures.

While Google Glass was preferred for visual feedback during a lecture, most preferred using a different device as input controls. Neither speech control nor head movements were seen as suitable in the context, and the touchpad at the side of Google Glass was seen as acceptable. However, regularly “scratching” one’s head to interact with the Google Glass touchpad was seen as a distraction. Given lecturers stand exposed in front of an audience, some worried that they could look ridiculous when interacting with Glass.

Desired information on Glass varied between the different participants and included (in no particular order) lecture notes, current slide, next slide, current time, countdown time, and titles of next slides. Other, more advanced features — like taking pictures, streaming video, or web searches — were seen as interesting, but not regarded as crucial for a first prototype nor used regularly in lectures.

3.2 Context Analysis

Preliminary informal tests with Google Glass revealed differences in display readability depending on the kind of room. Readability was better in flat seminar rooms as the image was usually seen against the background of a white wall. Lecture halls with rising rows of seats constitute a “noisy” background, thus reducing readability.

3.3 Conclusions from Analysis

Given the heterogeneity in desired information, even within the small sample size, a modular layout supporting different information types is needed. Given the “noisy”

background, especially in lecture halls, high contrast and clearly distinguishable colors are required. For controls, speech and head tracking is unsuitable for a presentation context. Either the touch pad of Glass has to be used, or the input has to be done via other devices like handheld remotes or smart watches.

4 User Interface Concept

How can the relevant information be presented for the lecturer on smart glasses? We look first at design guidelines and then present our prototype concept.

The Google Glass platform offers two different types of user interface patterns [Go17b]. The "timeline" pattern offers a consistent appearance for the home screens and settings of Google Glass. The "immersion" interface pattern is suitable for stand-alone applications that need to deviate from standard Google Glass applications, e.g. by implementing a custom user interface.

Given the use case — a specialized user interface for lecture support — the immersion type was chosen. The main screen of our user interface consists of three different areas. A three-row layout provides the heterogeneous target audience the flexibility to configure the layout to their preferences or the situational characteristics of the room by allowing for many different combinations of content modules. At the same time, it provides some structure to reduce information overload.

The main area in the center should handle information-heavy content, especially content modules displaying slides and notes. Two bars at the top and bottom are intended for additional information that does not take up much space, e.g., content modules displaying slide headings, timers or clocks. If areas are not used the other design elements scale up to use the otherwise unused space.

Fig. 2 shows an early mockup of the user interface. In this instance, the configured layout is comprised of six content modules:

- The top bar accommodates a module (1) showing heading of the current and the following three slides. It is modeled after announcement systems in public transport systems.
- In the center, the current slide (2) is shown on the left side, complemented with the matching presenter notes (3) on the right.
- The bar at the bottom contains three modules. The module on the left shows the time passed since the presentation was started (4). In the middle, a progress bar (5) displays the ratio of displayed to upcoming slides. The module on the right displays the current time of day (6).

A formative evaluation of the user interface concept revealed issues with readability

regarding the module in the top bar. Headings were frequently too long to display them for all four slides. Possible approaches for a solution included: scaling down font size, shortening headings or displaying fewer headings. However, each of these solutions poses new questions that needed further research. Thus, the particular module was removed in this first prototype.

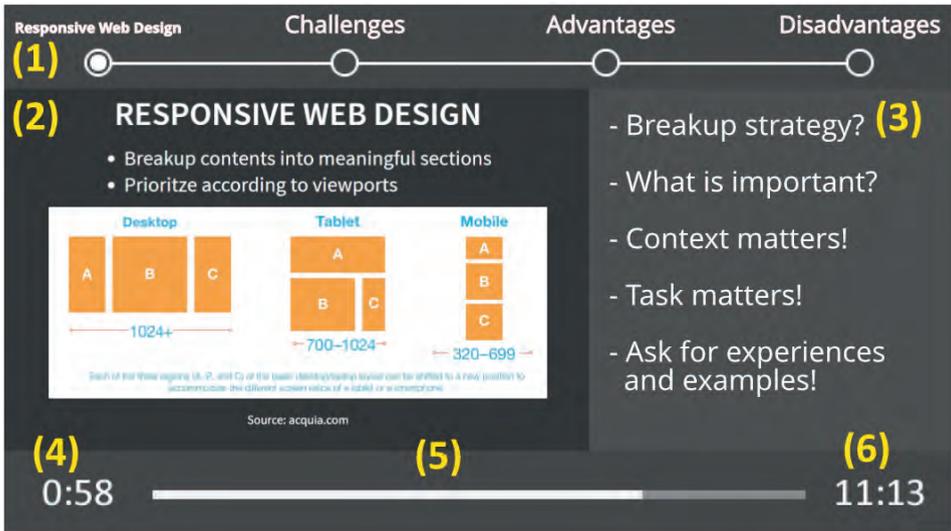


Fig. 2: Early user interface concept of the presentation view

The resulting user interface concept addresses both the heterogeneity of the user group and provides some structure to avoid information overload.

5 Backend System for Google Glass

A Google Glass application alone is not sufficient to realize the envisioned scenario. A backend system must be in place to control and deliver the lecture slides to the device. To keep development simple in this prototyping stage, the HTML presentation framework reveal.js [Re17] was chosen for presenting the slides. The use of HTML based presentations allows a clear separation between content and visual styling, enabling the use of different design templates on different devices. The devices were connected via WebSocket protocol by using the multiplexing plugin of reveal.js. The multiplexing plugin uses a Socket.io [So17] server and the publish-subscribe pattern to synchronize application events, e.g. display next slide, across connected devices. The developed system potentially allows for multiple devices being connected at the same time (including student notebooks and the like), but our focus is on a single lecturer.

The Google Glass App itself has been implemented using the Glass Development Kit

(GDK) with Android API level 19. The app consists, among others, of the following Android activities: In order to connect to the web-based presentation system, the app starts with the providing users with a Quick Response (QR) code. After successful connection, the *PresentationActivity* shows the previously chosen layout or a standard layout. The user can now navigate the slides of the loaded presentation by swiping horizontally on the touchpad. For this slide navigation, the interaction model from the Google Glass Timeline was adapted. Thus, a swipe from front to back moves the slide to the left, vice versa a swipe from back to the front moves the viewport to the right. Following the Google Glass design principles two other gestures have also been implemented: First, swiping down stops the App and brings the user back to the timeline. Second, a tap on the touchpad opens the *MenuActivity* of the app.

The menu is a substantial part of this application, as configuration and individualization of the user interface are very important both intra- and interindividually: First, as outlined above, the layout may be customized depending on user and physical context. Second, due to the current low availability of Google Glass, there might not be one device for every lecturer and switching layout profiles might be a regular task. To address these issues, possible solutions like configuration by a connected smartphone, the use of a web service or manual configuration were discussed. To keep the prototype simple, we opted for a manual approach. Settings relevant for a single presentation should be transferred via scanning of the QR code. More enduring settings, like the presentation layout, are configured in the menu.

Three different layouts were implemented for the summative evaluation of the prototype. One is shown in Fig. 3 primarily using the center region for displaying current and upcoming slides. The available slots in the top and bottom bar are not used, except for a timer in the bottom right displaying the elapsed time of the presentation.

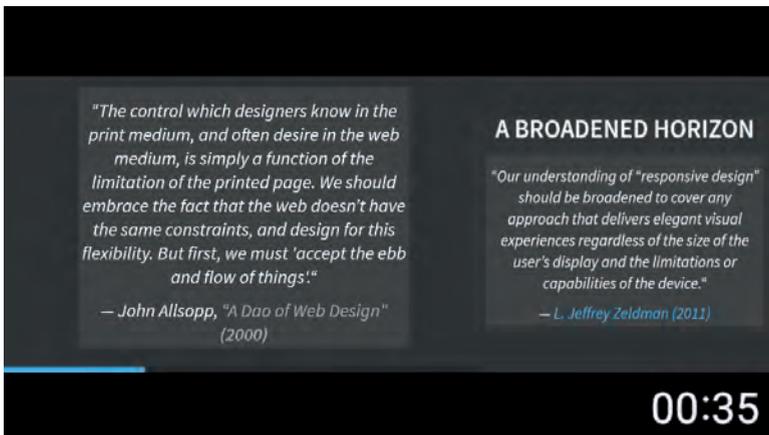


Fig. 3: One of the implemented layouts, showing current slide in the left, next slide on the right. The thin progress bar starting on the left center region is conveniently embedded in the

HTML presentation layout thus not taking any more limited display space. In the second layout option, the right main region showed the current display notes instead of previewing the next slide. The third layout variation makes use of dynamic column resizing by just showing the current slide in the center and adding a clock to the bottom bar. The color scheme of the slides is independent from the original layout template of the slide, allowing for high contrast.

This backend system provides the necessary infrastructure to allow lecturers to present with Google Glass and should be well-suited for a prototype test.

6 Summative Evaluation

To assess the quality of the developed prototype, a summative evaluation was conducted.

6.1 Method

The prototype was evaluated with six participants, four of which also participated in the user analysis. A structured system exploration with think aloud (cf. [SBS94]), followed by an interview was used. The evaluation was conducted in an open-space multimedia lab. Participants were asked to familiarize themselves with the touchpad and app navigation and try out different layouts with different color schemes (white on black and vice versa).

6.2 Results

Results of the summative evaluation include the following:

Display quality: Participants answers varied regarding the specific elements of the interface. Headers were readable, and the slide layout and the images were recognizable. Regarding the amount of text, beyond a few bullet points the text became hard to impossible to read. However, participants were able to recognize the specific slide even without being able to read the text and they were content with this information. The presenter notes were readable if they were concise. Similarly, the timer was easily readable. Varifocals pose a problem given the prism of Google Glass lies in the area of far-away objects of the varifocal glasses. Bright, direct light was also an issue, impairing readability.

Layout: While four participants were content with the layout, suggested improvements include additional variations, e.g. resizing modules or displaying only one module and clearly visible dividing lines between modules. The proportions of the modules were seen as well-chosen, especially the division of the central area into unequal slots.

Interaction: The touchpad was seen as the best way to interact with Google Glass on its

own. However, two participants felt it was odd and distracting to touch their head regularly for navigation, echoing the concerns of the analysis. Additionally, users got confused about the direction of the swipe movements. As mentioned above and following Google Glass conventions, to move from right to left (advance a slide) a swipe from front to back on the touchpad was chosen. Intuitively, users preferred it the other way around. Myo (armband for gesture interaction, [Myl7]) and smart watches were suggested as alternative control devices. Most would prefer to use the keyboard or a remote to control the slides. The optional head nudge feature was enabled to be able to switch off the display momentarily, which led to two participants switching off their display by accidental nodding.

Acceptance by students: Concerns were raised regarding acceptance by students. In addition to the scratch gesture of advancing slides, looking at the display was considered as quite noticeable and as potentially distracting for the audience.

Technical issues: While the delay between touch event and feedback on the projected slides was seen as acceptable, some network connection dropouts occurred.

Overall assessment: Participants did not see the technology as suitable for day-to-day lectures yet, but were open to additional tests and improvements and curious to see how the technology evolves.

6.3 Discussion

Overall, the summative evaluation did show the potential of Glass, but also highlighted important critical issues.

On the technological side, the current state of Google Glass poses hardware issues that may disrupt lecturers in their presentations. Among these are heat generation and corresponding sluggishness of the system, displaying text close to the border of the display, and unstable Internet connection, replicating some of the findings of [EME16].

The evaluation also highlighted some usability issues and minor errors that should be fixed in a next iteration. Users could be asked for their preferred swipe direction for navigating through the slides. The default control should be switched in accordance with the more intuitive movement. The metaphor of pushing the next slide to the students in front of the lecturer (moving the finger from back to front) vs. getting the slide back (moving from front to back) might be used to explain the slide controls. Variable color schemes and font sizes might solve some of the issues with readability. To deal with limited resolution automated scrolling concept might be a direction worth to explore.

Independent from the application, some users experienced headaches after wearing Google Glass for a few minutes. The causes were not explored in this study. Although one could argue that users might adapt to the small display over time, this issue has to kept in mind.

Regarding the range of features, the evaluation supported the outcomes of the initial user and task analysis, and lead to suggestions for improvement. Suggestions for content types can be easily implemented as the modular architecture allows to integrate new layouts or new content modules without developing a whole new application.

7 Conclusion and Future Work

In the present paper, we described a human-centered design process to assess whether lecturers can be supported with Google Glass and to identify the strengths and weaknesses of Google Glass in the lecture setting.

Our analysis did show the heterogeneity of users and their differing expectations, which was addressed in the modular user interface concept and backend system. Different modules can be used to personalize the graphical user interface. The backend system allows the flexible presentation of the content and allows for further adaptations and functions, e.g. providing the slides to students. The summative evaluation did test the developed prototype and provided empirical data on the success of the HCD.

Overall, results show that Google Glass has the potential to support lecturers. The participants were willing to try out the Google Glass App further, albeit only after further improvements have been made. However, the HCD process also revealed the difficulty of implementing a lecturer support system using smart glasses. Issues include major weaknesses of current technology (stability, heat, etc.) but also concerning unobtrusive interaction. Glass is useful to display information unobtrusively, but not for interacting unobtrusively with the device. Smartwatches might provide an alternative and less obtrusive way to control presentations. Also, smart glasses need to address the heterogeneity of the users. Lecturers differ in the kind of information they need and how it should be presented, thus care should be taken to maintain the customization and personalization features of the app.

The architecture used in the prototype is already capable to deliver the same content in different layouts/designs to the devices involved, e.g. one version intended for projecting onto canvas and a high contrast version to display on Google Glass. Ideally, this relieves lecturers from investing additional time into reformatting their slides for Google Glass. However, an integration of Google Glass with an authoring tool would give users more control over slide design and layout customizations. Developments will likely focus on integration with desktop presentation tools like Microsoft PowerPoint or Apple Keynote, as many users are familiar with them and have already invested lots of time and effort in their lectures using these applications.

An advantage of the proposed web-based solution is its expandability. This service can provide additional options to improve student interaction. For example, lecturers might allow access to a copy of their presentation on the personal devices of the students, e.g. smartphones or laptops. This would enable students to go a slide back in case they have

questions about the lecture. Incorporating smartphones or laptops opens up more use cases for students worth exploring, like annotating lecture slides, sharing content or take part in audience response activities. However, these are beyond the scope of this paper.

A major unknown at the moment is the perception of such a solution from the perspective of the students. In analysis and evaluation, lecturers were concerned about the impression they make on their students when using Google Glass. Thus, tests in actual lecture settings should include student evaluations.

Overall, the potentials of Google Glass are enticing. However, the technical hurdles – mostly related to the smart glass system itself – are currently too high for practical use. With the continued development of smart glasses these issues might soon be solved. We have also identified other areas in which more progress is needed and have identified ways to deliver and display content. Thus, contributing to the goal of allowing the lecturers to move more freely while staying on track and in close interaction with the students.

Acknowledgements

We thank the students and lecturers who took part in the user study and prototype evaluations for their valuable feedback.

Bibliography

- [EME16] Ebner, M.; Mühlburger, H.; Ebner, M.: Google Glass in Face-to-face Lectures - Prototype and First Experiences. *International Journal of Interactive Mobile Technologies*, 10/1, 2016.
- [Go17a] Google Glass Help: Tech specs, <https://support.google.com/glass/answer/3064128>, Stand: 01.03.2017.
- [Go17b] Google Developers: Glass UI, <https://developers.google.com/glass/design/ui>, Stand: 01.03.2017.
- [Hu15] Hu, G.; Chen, L.; Okerlund, J.; Shaer, O.: Exploring the use of google glass in wet laboratories. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, S. 2103–2108, 2015.
- [Ma16] Mason, M.: The MIT Museum Glassware Prototype: Visitor Experience Exploration for Designing Smart Glasses. *Journal on Computing and Cultural Heritage (JOCCH)*, 9/3, S.12, 2016.
- [MWH15] Mentler T.; Wolters C.; Herczeg M: Use cases and usability challenges for head-mounted displays in healthcare. *Current Directions in Biomedical Engineering*, 1/1, S. 534–537, 2015.
- [My17] Myo Gesture Control Armband, <https://www.myo.com/>, Stand: 01.03.2017.

- [Pr11] Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. DIN EN ISO 9241-210. Ergonomie der Mensch-System-Interaktion – Teil 210. Beuth, Berlin, 2011.
- [Re17] reveal.js – The HTML Presentation Framework, <http://lab.hakim.se/reveal-js/>, Stand: 01.03.2017.
- [SBS94] Someren, M. V.; Barnard, Y. F.; Sandberg, J. A.: The think aloud method: a practical approach to modelling cognitive processes. Academic Press, 1994.
- [So17] Socket.IO, <https://socket.io/>, Stand: 01.03.2017.
- [St13] Starner, T.: Project Glass: An Extension of the Self. IEEE Pervasive Computing, 12/2, S. 14–16, 2013.
- [Su68] Sutherland, I. E.: Head-mounted three-dimensional display. In: AFIPS '68 (Fall, part I): Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I. ACM, New York, S. 757-764, 1968.
- [We14] Weppner, J.; Hirth, M.; Kuhn, J.; Lukowicz, P.: Physics Education with Google Glass gPhysics Experiment App. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, New York, S. 279–282, 2014.

DIAL – Ein BigBlueButton-basiertes System für interaktive Live-Übertragungen von Vorlesungen

Luise Kaufmann¹, Tobias Welz² und Andreas Thor³

Abstract: Dieser Beitrag präsentiert **DIAL** (**D**istributed **I**nter**A**ctive **L**ecture), ein BigBlueButton-basiertes System für interaktive Live-Übertragungen von Vorlesungen. DIAL erweitert das Konferenzsystem BigBlueButton derart, dass Studierenden Zugang zum Chat- und Umfragesystem aus dem Stand-By ihres Endgerätes heraus ermöglicht wird, d.h. ohne die Notwendigkeit einer permanenten Verbindung. Damit ermöglicht DIAL Teilnehmern einer Live-Übertragung die einfache und direkte Interaktion analog zum virtuellen Klassenzimmer. Der Beitrag beschreibt die Architektur und die prototypische Implementation von DIAL und präsentiert Evaluationsergebnisse aus der praktischen Anwendung an der Hochschule für Telekommunikation Leipzig.

Keywords: E-Learning, Live-Übertragung, Virtuelles Klassenzimmer, BigBlueButton, Interaktion, verteilte Vorlesung

1 Motivation

Die Live-Übertragung von Vorlesungen per Webkonferenz ist ein weit verbreitet eingesetztes Mittel [et16], um Studierenden an unterschiedlichen Orten (z.B. Hörsäle der gleichen oder einer anderen Hochschule) die gemeinsame Teilnahme an derselben Lehrveranstaltung zu ermöglichen. Hochschulen reagieren damit auf die immer größere Zahl an Studierenden⁴ und die damit einhergehenden Ressourcenengpässe. Bauliche und räumliche Beschränkungen der Hochschulgebäude („Platzprobleme“) erlauben es bei teilnehmerstarken Studiengängen nicht mehr, dass alle Studierenden physisch an derselben Lehrveranstaltung teilnehmen können. Eine Aufteilung von Studiengruppen, bei der Dozierende ihre Lehrveranstaltungen mehrfach halten, stößt schnell an die Grenzen des Lehrdeputats der jeweiligen Dozierenden und wirft durch die Vergrößerung der Lehrveranstaltungszahl weitere Probleme hinsichtlich der Belegung der Hörsäle auf.

Gleichzeitig ermöglicht die Live-Übertragung (neben anderen E-Teaching-Aktivitäten, wie Vorlesungsaufzeichnungen etc.) Kooperationen zwischen Hochschulen bzw. Standorten. Wenn zuvor mehrfach gehaltene Lehrveranstaltungen nur noch einmalig an einer Hochschule bzw. an einem Standort durchgeführt werden, können die resultierenden Synergieeffekte zu finanziellen Entlastungen führen. Zusätzlich ermöglichen Live-Übertragungen die Erweiterung des Curriculums für Studierende, die nun an Vorlesun-

¹ T-Systems Int. GmbH, Telekom Security, Georg-Elser-Str. 8, 90441 Nürnberg, luise.kaufmann@telekom.de

² Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Str. 43-45, 04277 Leipzig, welz@hft-leipzig.de

³ Hochschule für Telekommunikation Leipzig, Gustav-Freytag-Str. 43-45, 04277 Leipzig, thor@hft-leipzig.de

⁴ <https://de.statista.com/statistik/daten/studie/221/umfrage/anzahl-der-studenten-an-deutschen-hochschulen>

gen teilnehmen können, die zuvor an ihrer Hochschule nicht angeboten wurden.

Die Teilnahme an einer live übertragenden Vorlesung stellt für die Studierenden keinen zusätzlichen Aufwand dar. Sie nimmt ihnen aber – im Vergleich zur Präsenzvorlesung – Möglichkeiten zur Interaktion, da Studierende keinen eigenen Kommunikationskanal zum Dozenten haben [Eb14, Kr16]. Fragen können in der Regel nur „moderiert“ gestellt werden, bei dem ein Moderator im entfernten Standort die Frage entgegennimmt und zum Dozenten asynchron (z.B. via Textchat) oder synchron (in Absprache mit den anderen Standorten) weiterleitet. Im Gegensatz dazu bieten virtuelle Klassenzimmer, bei denen Studierende über ein Webkonferenzsystem (z.B. Adobe Connect oder WebEx) an einer Lehrveranstaltung (Webkonferenz) teilnehmen [FH12], vielfältige Interaktionsmöglichkeiten, wie z.B. einen Textchat und die Teilnahme an Umfragen. Zusätzlich können Studierende über Mikrofon (und Webcam) ihre Fragen direkt an den Dozenten stellen.

Der Einsatz eines Webkonferenzsystems für die Live-Übertragung scheitert in der Regel an dessen höheren technischen Anforderungen in einem Hörsaal. Würde sich jeder Studierende mit seinem Endgerät (Notebook, Tablet etc.) im Hörsaal in die Webkonferenz einwählen, sendet der Server Medienstreams (Audio, Video, Whiteboard) an alle Studierenden, was eine sehr große Beanspruchung des Hochschul-WLANs inklusive zu erwartender Aussetzer bzw. Ausfälle zur Folge hat. Ein Verbindungsabbruch zwingt in der Regel zur Neuansmeldung in der Webkonferenz und stellt eine unnötige Ablenkung für die Studierenden dar. Ein weiterer Nachteil ist der hohe Stromverbrauch der Endgeräte beim dauerhaften Medienstreaming, der viele Studierende zwingen würde, ihre Geräte mit Strom zu versorgen, was in vielen Hörsälen nicht für alle Teilnehmer möglich ist.

Dieser Beitrag präsentiert DIAL, ein BigBlueButton-basiertes System für interaktive Live-Übertragungen von Vorlesungen. Dazu wurde das Open-Source Webkonferenzsystem BigBlueButton so erweitert, dass Medienstreams von der Nutzerinteraktion via Textchat und Umfragen getrennt verfügbar sind. Damit genügt es, dass nur ein Endgerät pro Standort (Hörsaal) als „normaler“ Teilnehmer der Webkonferenz beitrifft und die Medienstreams per Beamer und Lautsprecher in den Hörsaal verteilt werden. Studierende greifen aus dem Stand-By heraus mit einer eigens entwickelten App auf die Textchat- und Umfragefunktionalitäten zu und können so jederzeit in der Vorlesung interagieren. DIAL kombiniert damit die Vorteile einer Live-Übertragung (geringe Anforderung an Studierende) mit dem virtuellen Klassenzimmer (Interaktionsmöglichkeiten).

Das nachfolgende Kapitel beschreibt die Architektur von DIAL sowie die prototypische Implementation unter Verwendung von BigBlueButton. Kapitel 3 präsentiert eine kurze Evaluation und berichtet über die gemachten Erfahrungen mit DIAL an der Hochschule für Telekommunikation Leipzig (HfTL).

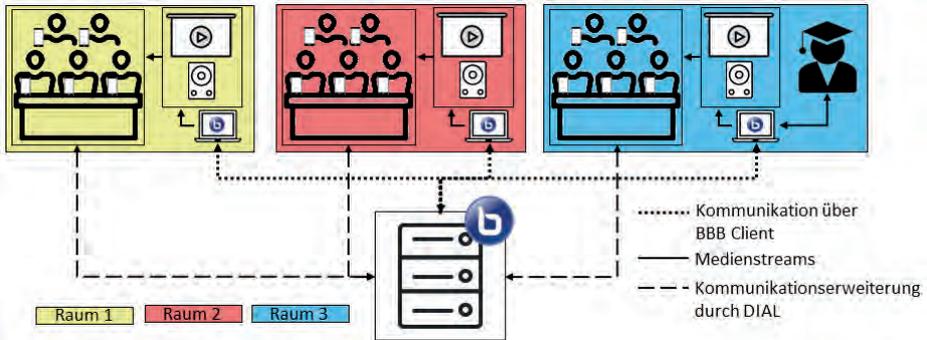


Abb. 1: Schematische Darstellung einer Live-Übertragung einer Vorlesung mit DIAL

2 Architektur und Implementation

Das prinzipielle Schema einer Live-Übertragung einer Vorlesung sieht vor, dass sich die Teilnehmergruppen in mehreren Räumen verteilt befinden (siehe Abb. 1). DIAL basiert dazu auf einem Webkonferenzsystem, welches die Verarbeitung der Medienstreams (Audio- und Videosignale) sowie die Kommunikation aller Teilnehmer steuert. In jedem Veranstaltungsraum tritt ein technischer Supportmitarbeiter der initiierten Konferenzsitzung bei und streamt die übertragenen Sitzungsdaten über einen Beamer und Lautsprecher zum Auditorium. Die Nutzer im Veranstaltungsraum hören den Audiostrom der Dozierenden und können über die Projektion das Whiteboard, die Präsentationsfolien, die geteilten Anwendungen, den Chat, das Videobild der Dozierenden und die Ergebnisse von Umfragen sehen. Ob und wann diese Inhalte geteilt werden, entscheiden die Dozierenden, die der Sitzung mit Moderationsrechten beitreten. Die Lehrveranstaltung wird entweder in einem der Veranstaltungsräume oder dezentral gehalten.

Damit eine Übertragung von Informationen von Studierenden zu Dozierenden stattfindet, wird die Kommunikation durch DIAL erweitert. Hierbei nutzen die Studierenden im Auditorium ihre privaten Endgeräte (z.B. Smartphones oder Tablets), um Chatnachrichten zu lesen bzw. zu schreiben oder an Umfragen teilzunehmen. Diese Nutzerinteraktionen umfassen i. Allg. nur wenige Kilobyte und belasten das Mobilfunknetz oder Hochschul-WLAN nur gering. Im Vorfeld durchgeführte Skalierungs- und Lasttests mit simulierten Nutzen ließen außerdem erkennen, dass eine Teilnahme mehrerer Hundert Studierender mit vorhandener Infrastruktur problemlos möglich ist (vgl. [Ka16] Kapitel 7).

DIAL wurde auf Basis des Open-Source Webkonferenzsystems BigBlueButton entwickelt, da dieses bereits für Teletutorien an der HfTL regelmäßig verwendet wird, leicht erweiterbar ist und sich in die weit verbreiteten Lernplattformen Moodle und ILIAS nahtlos integriert. Auch bietet es vielfältige Funktionen, u.a. das automatische Generieren einer Umfrage an Hand einer Präsentationsfolie, bei dem die Antwortmöglichkeiten

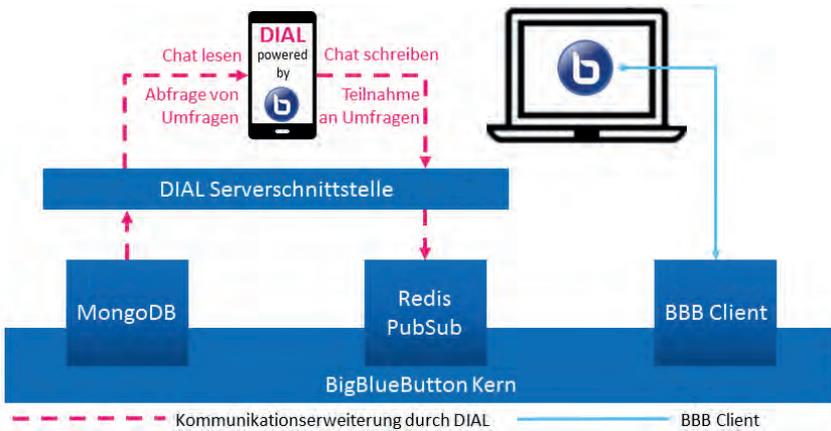


Abb. 2: Informationsfluss zwischen BigBlueButton-Kern und DIAL-App (Smartphone eines Studenten; links) sowie regulärem BigBlueButton-Client (Notebook des Dozenten; rechts)

automatisch von der Folie extrahiert werden.

Die Dozenten- und Clientrechner zur Hörsaalübertragung (siehe Notebook in Abb. 2) betreten die Veranstaltung über einen der mitgelieferten Standard-Clients. Von den BigBlueButton-Entwicklern wird über GitHub der komplette BigBlueButton-Kern (vgl. Abb. 2) zur Verfügung gestellt, der verschiedene Schnittstellen und Module enthält.⁵ Als zentraler Kommunikationskanal zwischen den Modulen dient die Komponente *Redis PubSub*, welche das Lesen und Schreiben von Nachrichten in verschiedenen Channels steuert. Eine *MongoDB*-Datenbank dient als zentraler Speicherort für Sitzungsinformationen, wie z.B. den kompletten Textchat und alle Umfragedaten.

Die serverseitige DIAL-Schnittstelle fungiert nun als Vermittler zwischen den mobilen Clients (Smartphones der Studierenden) und dem BigBlueButton-Kern. Sie stellt die folgenden Funktionen zur Verfügung: *Chat lesen*, *Chat schreiben*, *Abfrage von Umfragen* und *Teilnahme an Umfragen*. Dazu liest sie die erforderlichen Sitzungsinformationen aus der *MongoDB* aus und ist in der Lage, entsprechende Channel-Nachrichten in der definierten Syntax in den *Redis PubSub* zu übermitteln bzw. auszulesen. Die DIAL-Schnittstelle funktioniert daher unabhängig von Änderungen am BigBlueButton-Kern, solange das Nachrichtenformat des *Redis PubSub* und die Speicherung der Sitzungsdaten in der *MongoDB* gleich bleiben.

Für die mobilen Clients wurde eine PHP-basierte Webapplikation entwickelt, welche über eine ansprechende GUI die entsprechenden Funktionen der DIAL-Schnittstelle den Nutzern zugänglich macht. Damit benötigen die DIAL-Clients (siehe Smartphone in Abb. 2) lediglich einen Webbrowser und keine zusätzliche Software.

⁵ <http://docs.bigbluebutton.org/>

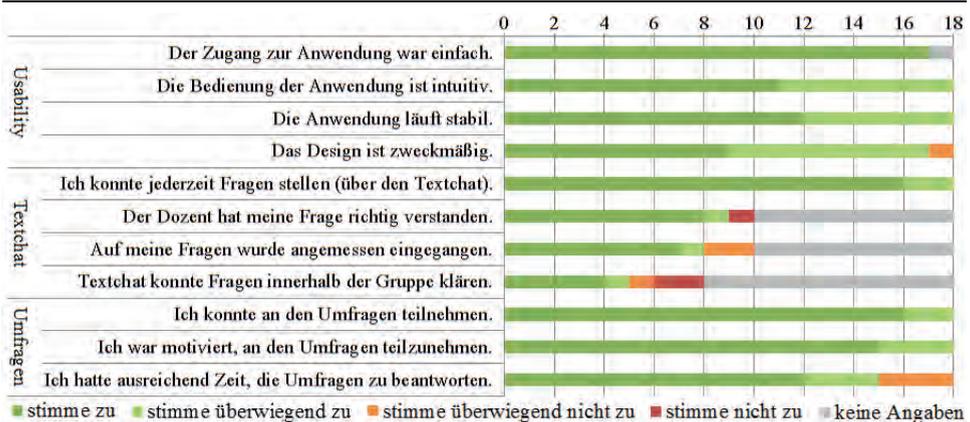


Abb. 3: Ergebnisse des formativen Feedbacks der Teilnehmer der Probevorlesung

3 Use Case und Evaluation

DIAL wurde im Rahmen einer Probevorlesung an der HfTL mit insgesamt 20 Studierenden in zwei Räumen eingesetzt. Abb. 3 zeigt die Zusammenfassung des formativen Feedbacks der Teilnehmer, die zusätzlich in einer Gruppendiskussion ihre Erfahrungen äußerten. Sowohl die Webanwendung als auch BigBlueButton mit der DIAL-Erweiterung liefen fehlerfrei und konnten von Studierenden und dem Dozenten problemlos genutzt werden. Die Analyse der Auslastung (Datenverkehr, Server-Load) bestätigte die deutlich niedrigeren Anforderungen im Vergleich zur normalen Nutzung eines Webkonferenzsystems (siehe auch [Ka16] für Details und weitere Evaluationsergebnisse).

Wie üblich bei Live-Übertragungen stellt der (im Vergleich zur normalen Vorlesung) technische Aufwand hohe Ansprüche an den Dozenten und ggf. das unterstützende Personal. Besonders nachteilig ist aus Sicht des Dozenten das fehlende non-verbale Feedback der Studierenden in den entfernten Räumen. Dies unterstreicht noch einmal die Wichtigkeit der Kommunikationsmöglichkeit zwischen (allen) Studierenden und dem Dozenten. Die Studierenden nutzten daher den Textchat der Webanwendungen vornehmlich für Fragen an den Dozenten, sahen aber Verbesserungsbedarf bezüglich deren Beantwortung. Die verzögerte Beantwortung sowie die fehlende Rückkopplung, ob der Dozent die Frage bemerkt und richtig verstanden hat, wurden kritisiert. Da der Textchat für alle Studierenden lesbar war, konnten einige Fragen bereits in Kleingruppen innerhalb der Studierendengruppe geklärt werden, was wiederum dem Dozenten entging.

Besonders positiv wurden die Umfragen von den Studierenden bewertet, die aus Sicht des Dozenten gleichzeitig zwei Funktionen hatten. Primär erlaubten sie dem Dozenten Fragen (z.B. in Single-Choice-Form) an alle Teilnehmer zu stellen, um das Verständnis der Studierenden für den weiteren Verlauf der Vorlesung zu sichern. Gleichzeitig konnte der Dozent die Zeit, während die Studierenden die Frage beantworten, nutzen, um den

Textchat zu lesen und im Anschluss ggf. darauf einzugehen. Das Feedback der Studierenden zeigt auch, dass sie tendenziell mehr Zeit zur Beantwortung der Fragen wünschen, was dem Dozenten wiederum mehr Zeit zum Lesen gibt. Dozenten sollten also bei dieser Art der Vorlesung ihr didaktisches Konzept so anpassen, dass in regelmäßigen Abständen Umfragen möglich sind und ggf. auch kommunizieren, dass nach Umfragen auch auf Fragen des Textchats eingegangen werden kann, um so die Erwartungshaltung der Studierenden auszurichten. Insgesamt wurde die Live-Übertragung mit DIAL als sehr positiv bewertet. Um die Kommunikation zwischen Studierenden und Dozenten weiter zu verbessern, wünschten sich die Teilnehmer weitere Formen des strukturierten Feedbacks über die Webanwendung geben zu können, z.B. Emoticons für eine Handmeldung oder einen Hinweis, dass die Vorlesungsgeschwindigkeit zu schnell wird.

4 Zusammenfassung und Ausblick

Dieser Beitrag stellte Architektur und prototypische Implementation von DIAL vor, einem System für interaktive Live-Übertragungen von Vorlesungen. Dazu wurde das Webkonferenzsystem BigBlueButton um eine Schnittstelle erweitert, die den Veranstaltungsteilnehmern Zugang zum Chat- und Umfragesystem ermöglicht. Die DIAL-Architektur realisiert dabei eine ressourcenschonende Kommunikation zwischen den mobilen Endgeräten der Teilnehmer und dem Konferenzsystem, um eine große Einsatzfähigkeit auch bei begrenzter Hörsaalinfrastruktur (Netzwerk, Steckdosen) zu ermöglichen. Ein Praxistest im Rahmen einer Vorlesung an der HfTL verlief positiv und zeigte das Potenzial von DIAL. Zukünftige Arbeiten erweitern die DIAL-App um zusätzliche Möglichkeiten der Interaktion und untersuchen, in welchen didaktischen Szenarien DIAL Vorteile gegenüber klassischen Live-Übertragungen bzw. Webkonferenzen bietet.

Literaturverzeichnis

- [Eb14] Ebner, M.; Haintz, C.; Pichler, K.; Schön, S.: Technologiegestützte Echtzeit-Interaktion in Massenvorlesungen im Hörsaal. Entwicklung und Erprobung eines digitalen Backchannels während der Vorlesung. In (Rummler, Klaus, Hrsg.): Lernräume gestalten - Bildungskontexte vielfältig denken. Waxmann, 2014.
- [et16] e-teaching.org „Liveübertragung“, <https://www.e-teaching.org/lehrszenarien/vorlesung/liveuebertragung>, Stand: 03.02.2017.
- [FH12] Ferdinand, P.; Heckmann, P.: CSCL-Kompendium 2.0: Lehr- und Handbuch zum computerunterstützten kooperativen Lernen. Oldenbourg, Kapitel Plattformen, 2012.
- [Ka16] Kaufmann, L.: Entwicklung einer BigBlueButton-basierten Anwendung zur Interaktion zwischen Dozent und Studierenden innerhalb dezentral gehaltener Vorlesungen. Bachelorarbeit, Hochschule für Telekommunikation Leipzig, 2016.
- [Kr16] Krauss León, C.: Interaktion: Live-Feedback-Systeme und Evaluationstools in Vorlesungen. Masterarbeit, Technische Universität Dresden, 2016.

Seamless Learning in Lake Constance Region

Christian Rapp¹, Bernadette Dilger², Ralf Schimkat³, and Rainer Mueller³

Abstract: As part of large-scale project in the Lake Constance region that borders Germany, Austria and Switzerland, seamless learning prototypes are designed, implemented, and evaluated. A design-based research (DBR) approach is used for the development, implementation and evaluation. The project consists of one base project that brings together instructional design and instructional technology specialists as well as (educational) software architects, supporting seven subprojects in developing seamless-learning lighthouse implementations mostly within higher and further education. Industrial partners are involved in all subprojects. The authors are unaware of similar seamless learning projects using a DBR approach that include evaluation and redesign and allow for comparison of experiences by applying DBR across subprojects. The project aims to generate novel seamless learning implementations, respective novel research and novel software supporting different aspects of seamless learning. The poster will present the project design, delineate areas of research and development, and include initial empirical results.

Keywords: seamless learning, design based research, design principles, educational design theory, instructional technology.

Introduction

Kuh [Ku96] coined the term seamless learning (SL), initiating the first phase of SL. The widespread and affordable availability of mobile devices with ubiquitous internet access fostered a second, more technically orientated, phase of SL. Wong [Wo15] points to the lowest common denominator of the two phases: “the continuity of individual learners’ experiences across multiple learning spaces, particularly to connect formal and informal learning spaces” (p. 5). Wong also reviews the development of the research field and synthesizes existing definitions: “Seamless learning is when a person experiences a continuity of learning, *and consciously bridges the multifaceted learning efforts*, across a combination of locations, times, technologies or social settings” [Wo15, p. 10]. Consequently, SL draws our attention to the various *contexts* within which learning takes place, reveals respective requirements and asks how to align them in order to allow for SL *across* contexts (e.g., formal and informal learning, personal and group learning, physical and digital etc.). A focus on the transition from one educational context to another (e.g. from secondary to tertiary or further education) adds the lifelong learning

¹ Zurich University of Applied Sciences, Center for Innovative Teaching and Learning, School of Management and Law, St. Georgenplatz 2, Winterthur, Switzerland, 8400, rapp@zhaw.ch

² University of St. Gallen, Switzerland, Institute for Business Education and Educational Management, Dufourstrasse 40a, 9000 St. Gallen, Switzerland, bernadette.dilger@unisg.ch

³ Hochschule Konstanz University of Applied Sciences, Department of Computer Science, Alfred-Wachtel-Straße 8, 78462 Konstanz, Germany, ralf.schimkat@htwg-konstanz.de, rainer.mueller@htwg-konstanz.de.

perspective to SL. A central aim of the project is to identify gaps between various context requirements with the various stakeholders of the subprojects.

Design-based research (DBR) will be employed to simultaneously develop practical, relevant learning designs and empirically-tested design principles for SL. Within the DBR approach, a cyclical process will be followed for problem understanding, design, testing, evaluation and redesign in each of the disparate subprojects [MR12]: During workshops, the authentic contexts of learners' and teachers' experiences will be explored, resulting in requirements drawn up from the identified gaps of the subprojects. These requirements will be synthesised to form a problem landscape showing the different gaps, respective explanatory factors, first design heuristics and design ideas. The agile process of prototyping, testing and evaluation iterations will help to directly understand the impact of different design ideas and give evidence to effective design approaches. Through ongoing formative evaluation, empirical data will be collected on learners' experiences, as well as the outcomes and effectiveness of the design. Through redesign, the prototype will be further developed in different stages and the design improved until sufficiently robust with a context-sensitive theory of SL [De03, Eu14].

As an integrated SL approach, the technological subproject supervising of suitable tools and system support will become an indispensable part of the project. Experiences and identified media gaps gathered in testing and evaluation will optimise tool architecture and best-of-breed functionalities chosen for the next iterations, thus resulting in a more and more elaborated platform as general purpose SL technological-decision-support tool. Over a period of four years, not only will each subproject undergo multiple iterations, but the development of the most suitable practice solutions and design theories will be fostered by staggering subprojects to make best use of prior experiences to enhance subsequent development. Systematic comparison between subprojects aims to overcome the idiosyncratic perspective of a single specific design/principle; crucial for DBR.

References

- [De03] Design Based Research Collective (2003): Design-Based Research: an emerging paradigm for educational inquiry. *Educational Researcher* 32/01, S. 5-8, 2003.
- [Eu14] Euler, Dieter: Design Research – A Paradigm under Development. In (Euler, D.; Sloane, P.F.E., Hrsg.): *Design-Based Research, Zeitschrift für Berufs- und Wirtschaftspädagogik* 27(Beiheft), Stuttgart, S. 15-41, 2014.
- [Ku96] Kuh, George: Guiding principles for creating seamless learning environments for undergraduates. *College Student Development*, 37/02, S. 135-148, 1996.
- [MR12] McKenney, Susan; Reeves, Thomas: *Conducting educational design research*, Routledge, New York, 2012.
- [Wo15] Wong, Lung-Hsiang: A Brief History of Mobile Seamless Learning. In (Wong, L-H.; Milrad, M.; Specht, M., Hrsg.): *Seamless Learning in the Age of Mobile Connectivity*, Springer, Singapore, 2015.

Das E-Portfolio in der Kölner LehrerInnenbildung: Wie kann durch das elektronische Portfolio in ILIAS die kontinuierliche Selbstreflexion von Lehramtsstudierenden an der Universität zu Köln unterstützt werden?

Henning Host¹, Svenja Jaster², Dominik Baedorf³ und Maria Boos⁴

Abstract: Das digitale Poster wird in Form eines monoskopischen 360°-Videos präsentiert. Themen des 360°-Videos sind ILIAS als Learning-Management-System, das „Portfolio Praxisphasen“ als Reflexionsinstrument und das E-Portfolio in der Kölner LehrerInnenbildung. Die Rolle als „reflective practitioner“ wird mit dem Portfolioansatz in der Kölner LehrerInnenbildung aufgebaut und gefördert. Das 360°-Video stellt dabei ein zusätzliches mögliches Reflexionsinstrument dar. Bereits in der Ausbildung werden Studierende mit diesem Poster für die Potentiale multimedialer und immersiver Welten sensibilisiert.

Keywords: 360°-Video, E-Portfolio, reflective practitioner, LMS, LehrerInnenbildung.

1 Darstellungsform

Das Poster wird als monoskopisches 360°-Video präsentiert. Ausgangspunkt des Videos ist ein digitales Poster zum E-Portfolio in der Kölner LehrerInnenbildung, von dem aus ILIAS als Learning-Management-System, das Portfolio als Reflexionsinstrument und das E-Portfolio in der Kölner LehrerInnenbildung vorgestellt und diskutiert wird. Das 360°-Video schafft ein immersives Erleben der Diskussion und des Austausches zum E-Portfolio, erzeugt eine stärkere Verbindung zwischen NutzerInnen und Inhalt und reagiert auf veränderte Sehgewohnheiten in Videoportalen [Pr13].

2 Hintergrund

Mit der Reform des Lehrerausbildungsgesetzes in Nordrhein-Westfalen von 2009 rückt der Berufsfeld- und Praxisbezug in den Fokus der LehrerInnenbildung [St16]. Im Sinne der „systematischen Dokumentation, Reflexion und dialogischen Steuerung des individuellen Kompetenzaufbaus“ [St16] wurde das obligatorische „Portfolio Praxiselemente“ eingeführt. Wesentliches Ziel der Kölner LehrerInnenbildung ist der Aufbau einer reflexiven Haltung. Die Rolle als „reflective practitioner“ [Sc83] wird mit dem Portfolioansatz gefördert.

¹ Zentrum für LehrerInnenbildung, Köln, Albertus-Magnus-Platz, 50923 Köln, HHost@uni-koeln.de

² Zentrum für LehrerInnenbildung, Köln, Albertus-Magnus-Platz, 50923 Köln, Svenja.Jaster@uni-koeln.de

³ Zentrum für LehrerInnenbildung, Köln, Albertus-Magnus-Platz, 50923 Köln, Dominik.Baedorf@uni-koeln.de

⁴ Zentrum für LehrerInnenbildung, Köln, Albertus-Magnus-Platz, 50923 Köln, Maria.Boos@uni-koeln.de

3 Das E-Portfolio

Das „Portfolio Praxisphasen“ in der Kölner LehrerInnenbildung wird als E-Portfolio im Learning-Management-System ILIAS geführt. Die Universität zu Köln ist damit einer von wenigen Standorten bundesweit, an denen das Portfolio rein elektronisch geführt wird. Die obligatorischen Bestandteile des Portfolios fokussieren insbesondere die Themen Motivation, Ziele und eignungsreflexive Fragestellungen. Die einzelnen obligatorischen Reflexionsbausteine der drei Praxisphasen sind einander angeglichen und bauen aufeinander auf. Die Gestaltung des Portfolios als elektronisches Portfolio bietet verschiedene Vorteile, wie z. B. die Integration aller Medienformate oder die Freigabe von Teilen oder ganzen Portfolios in ILIAS.

4 Mehrwert von 360°-Videos für das E-Portfolio

Mit der Aufzeichnung von Unterrichtssequenzen durch Studierende in Form von 360°-Videos wird eine neue Möglichkeit der Reflexion vorgestellt. Damit wird ein komplexer Einblick in Unterrichtsprozesse ermöglicht [TS10]. Das Verhalten der Lehrkraft sowie der Lernenden können aufgezeichnet und wiedergegeben werden. Die Studierenden können damit ihre professionelle Selbstwahrnehmung schärfen [TS10]. Insbesondere Diskussionsformate, wie z.B. Podiums- oder Plenumsdiskussionen, können durch 360°-Videos in ihrer Gesamtheit aufgezeichnet und für Reflexionsabläufe verwendet werden. Studierende werden weiterhin für die Potentiale multimedialer und immersiver Welten sensibilisiert.

Literaturverzeichnis

- [Pr13] Prager, G.: 360° als neues Lernmedium? Entwicklungen im Bildungsbereich. In (Overschmidt, G.; Schröder, V., Hrsg.): Fullspace-Projektion. Mit dem 360°lab zum Holodeck. Springer, Berlin, S. 213-221, 2013.
- [Sc83] Schön, D.: The Reflective Practitioner. How Professionals Think in Action, Basic Books, New York, 1983.
- [St16] Stiller, E.: Das „Portfolio Praxiselemente“ als personenorientierter Lernbegleiter durch die berufspraktische LehrerInnenausbildung in Nordrhein-Westfalen. In (Boos, M. et al. Hrsg.): Portfolioarbeit phasenübergreifend gestalten. Konzepte, Ideen und Anregungen aus der LehrerInnenbildung. Waxmann, Münster, S. 11-19, 2016.
- [TS10] Trautmann, M.; Sacher, J.: Videofeedback als Instrument zur Unterrichtsentwicklung. In (dies., Hrsg.): Unterrichtsentwicklung durch Videofeedback. Besser kommunizieren lernen. Vandenhoeck & Ruprecht, Göttingen, 11-24, 2010.

Modulares Lernen online: ein interaktiver Brückenkurs in das Geophysik-Masterstudium (iBRIDGE)

David Lohner¹

Abstract: Am Karlsruher Institut für Technologie (KIT) entsteht ein onlinebasierter Brückenkurs in das Geophysik-Masterstudium. In diesem Projekt werden mehrere didaktische Szenarien abgebildet und in einem OER-Repositorium ein digitaler Lehrmittelpool aufgebaut. Das Poster zu diesem Beitrag ist online abrufbar.²

Keywords: Onlinekurs, Lehrvideo, individuelle Lernpfade, OER

1 Hintergrund

In Masterstudiengängen ist das Wissensniveau der Studienanfänger, die teilweise aus ganz anderen – wenn auch angrenzenden – Fächern kommen, sehr unterschiedlich. Institute aller Fakultäten stehen vor der Herausforderung, diese Heterogenität mit möglichst geringem Ressourcenaufwand kompensieren zu müssen; vor allem die Institute der so genannten »Kleinen Fächer«^{3,4} mit wenig Lehrpersonal. Um diesem Umstand entgegenzuwirken, fördert das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg das Projekt iBRIDGE⁵ (interaktiver **BR**ückenkurs **In Das GE**ophysik-Masterstudium).

2 iBRIDGE Steckbrief

Im Masterstudium des Kleinen Fachs Geophysik am KIT werden auf Grund der Heterogenität des Vorwissens der Studierenden bereits in frühen Phasen des Studiengangs einzelne überfordert, während Absolventen des Bachelors Geophysik unterfordert sind. Beides kann dazu führen, dass Studierende ihr Studium abbrechen und so die Studierendenzahl in der Geophysik abnimmt.

Das Projekt iBRIDGE hat Modellcharakter für alle Masterstudiengänge insbesondere der Kleinen Fächer. Der Kurs trägt dazu bei, die Qualität der Lehre zu verbessern: Die Lehrveranstaltungen im Masterstudium können sich von Beginn an auf die relevanten

¹ Karlsruher Institut für Technologie (KIT), Zentrum für Mediales Lernen (ZML), Karl-Friedrich-Str. 17, 76133 Karlsruhe, david.lohner@kit.edu

² Poster online unter <http://bit.ly/delfi2017-lohner>

³ vgl. <http://www.kleinefaecher.de/in-abgrenzung-zu-grossen-faechem> [Abruf: 16.06.2017]

⁴ vgl. <http://www.kleinefaecher.de/in-abgrenzung-zu-spezialgebieten> [Abruf: 16.06.2017]

⁵ Mehr Infos: <http://www.zml.kit.edu/iBridge.php>

Schwerpunkte konzentrieren, ohne zunächst ein homogenes Wissensniveau der Studierenden herstellen zu müssen. So bleibt mehr Zeit für die eigentliche Wissensvermittlung und Vertiefung. Davon profitieren die Studierenden, da sie so höher qualifiziert werden.

Der Kurs besteht aus insgesamt fünf Modulen und wird von den Studierenden nach dem Abschluss eines Bachelorstudiengangs, aber vor Beginn des Masterstudiengangs durchlaufen. Die Module setzen sich zu großen Teilen aus Videoinhalten und Begleitmaterial in Form von Skripten, Literaturhinweisen und Übungsaufgaben zusammen. Es gibt zwei Fachmodule, die jeweils aus 20 Lehrvideos bestehen. Studierende können mit dem Begleitmaterial eigene Defizite aufarbeiten. In einem weiteren Modul aus zehn Lehrvideos geht es um Einführungen in gängige Software, die für geophysikalische Modellrechnungen herangezogen wird. Zusätzlich zu diesen drei spezifischen Modulen beinhalten zwei weitere jeweils Aufzeichnungen von Einführungsvorlesungen des Bachelorstudiengangs Geophysik. Diese Aufzeichnungen werden verschlagwortet und durchsuchbar gemacht. Sie können entweder als Vorlesung angesehen oder als videobasiertes Nachschlagewerk verwendet werden. Den Modulen jeweils voran- und nachgestellt sind Online-Tests, deren Ergebnisse sich in individuellen Empfehlungen niederschlagen, welche Inhalte die Defizite des Einzelnen kompensieren können.

3 Erweiterung des Projekts

Das Projekt besteht neben dem Onlinekurs aus einem weiteren wesentlichen Baustein: In Kooperation mit der Universität Tübingen wird im dort entstehenden zentralen OER⁶-Repositorium für die Hochschulen des Landes Baden-Württemberg ein digitaler Lehrmittelpool angelegt, der allen Geophysik-Standorten zu Gute kommen soll. Darüber hinaus werden die einzelnen Module des Kurses sukzessive ins Englische übersetzt: Dabei werden nur Untertitel und grafische Elemente übersetzt, sodass der Lernende zur deutschen Tonspur wahlweise deutsche oder englische Untertitel und deutsche oder englische Textelemente sehen kann. So wird der Transfer des Fachvokabulars in beide Richtungen gefördert, was bei einer internationalen Disziplin wie der Geophysik von Vorteil ist.

Außerdem bietet der Kurs Anknüpfungspunkte für weitere Forschung, die sich mit der Evaluation der (Lehr-)Qualität von Onlinekursen befasst⁷ oder Möglichkeiten der Vernetzung zwischen den einzelnen Geophysik-Standorten in Deutschland untersucht.

⁶ Open Educational Resources

⁷ Vgl. Lohner, D.: »Vergleichbare Bewertung verschiedener videobasierter Lernmodule am Beispiel der Geophysik« in diesem Band

Vergleichbare Bewertung verschiedener videobasierter Lernmodule am Beispiel der Geophysik

David Lohner¹

Abstract: Die voranschreitende Digitalisierung der Hochschullehre verlangt nach neuen Lehrformaten und deren adäquaten Evaluation. Im Rahmen eines Promotionsvorhabens wird am Karlsruher Institut für Technologie (KIT) ein Evaluationskonzept für einen Brückenkurs zwischen dem Bachelor- und Masterstudiengang Geophysik² in Anlehnung an Ansätze aus Design-Based Research entworfen und iterativ angepasst. Das Poster zu diesem Beitrag ist online abrufbar.³

Keywords: Onlinekurs, Lehrvideo, Design-Based Research, Evaluation

1 Überblick und Kontext

Die Digitalisierung hält in unterschiedlichen Ausprägungen Einzug in die universitäre Lehre. Das Medium »Webvideo« als Hauptvermittlungsträger hat sich dabei als Kernelement dieser Lehreinheiten durchgesetzt [Re16]. Lehrveranstaltungen, die bisher den klassischen Formaten Vorlesung und Seminar zuzuordnen waren, werden in videobasierte »Flipped Classrooms« oder »Blended Learning« Angebote verwandelt. Am Zentrum für Mediales Lernen (ZML) des KIT entstehen in vielfältigen Kontexten solche Onlinekurse. Unabhängig von der fachlichen Ausrichtung sind sie trotz verschiedener Anforderungen strukturell ähnlich aufgebaut. Für eine vergleichbare Evaluation dieser digitalen Lehrformate existiert am KIT derzeit kein geeignetes Werkzeug.

2 Evaluation des Onlinekurses

Auf Grund der verschiedenen Module sind in iBRIDGE² unterschiedliche didaktische Szenarien abgebildet. Das breite Spektrum didaktischer Gestaltung, das dem Kurs bereits innewohnt, bietet sich an, ein Forschungsvorhaben eng an das Projekt zu koppeln. Der Onlinekurs wird nach jeder Durchführung mit Fragebögen an die Studierenden analysiert. Dabei liegt das Hauptaugenmerk auf der Ausgestaltung der einzelnen Module, dem Zusammenwirken unterschiedlicher Medien (Video, Texte etc.) und den Querverweisen zwischen den Modulen.

¹ Karlsruher Institut für Technologie (KIT), Zentrum für Mediales Lernen (ZML), Karl-Friedrich-Str. 17, 76133 Karlsruhe, david.lohner@kit.edu

² Vgl. Lohner, D.: »Modulares Lernen online: ein interaktiver Brückenkurs in das Geophysik-Masterstudium (iBRIDGE)« in diesem Band

³ <http://bit.ly/delfi2017-lohner>

Ziel eines Promotionsvorhabens ist es, in einem iterativen Prozess Indikatoren auszumachen, die jeweils auf das betreffende Modul und seine didaktischen Eigenheiten zugeschnitten, und davon ausgehend auf andere Onlinekurse – auch anderer Fachrichtungen – übertragbar sind: Wie wichtig sind für die jeweiligen Formate technische Hilfestellungen in Form von auswählbaren Kapiteln, einer frei wählbaren Wiedergabegeschwindigkeit, Pausen oder Fragen innerhalb/nach den Videos (vgl. [MS14])? Gibt es dabei Unterschiede zwischen kurzen und langen Videos? Welchen Anforderungen muss das Zusatz- oder Begleitmaterial entsprechen, um den Lernprozess optimal zu unterstützen? Nach welchen Kriterien ist die Kommunikation während der Kursdurchführung zwischen den Studierenden untereinander und mit den Dozierenden zu bewerten? Darüber hinaus sollen anonymisierte statistische Daten der Videoplattform analysiert werden und diese mit den Auswertungen der Befragungen verknüpft werden. Auf diese Weise können Aussagen über die Qualität und Nutzung des eigentlichen Videomaterials gemacht werden.

Das Promotionsvorhaben lehnt sich an die Ansätze des Design-Based Research [Se15] an, da die iterative Vorgehensweise zum Entwicklungsprozess des Kurses passt. So kann das Lehrangebot durch Interventionen an die aus den Fragebögen gewonnenen Erkenntnisse angepasst und verbessert werden. In weiteren Interventionen werden die Anpassungen dann erneut evaluiert. Auf diese Weise entsteht sukzessive ein Kriterien- und Fragenkatalog, mit dem sich Onlinekurse je nach Ausprägung (abgebildet durch die verschiedenen Module) bewerten lassen. Durch diesen Katalog werden die Evaluationen untereinander vergleichbar. Die Erkenntnisse sollen nicht nur in ein Evaluationswerkzeug überführt werden, sondern auch dazu dienen, iBRIDGE selbst weiterzuentwickeln und künftig ähnliche Kurse für andere (Kleine) Fächer zu entwerfen und durchzuführen.

Literaturverzeichnis

- [MS14] Merkt, M.; Schwan, S.: Training the use of interactive videos: effects on mastering different tasks. *Instructional Science* 42/3, S. 421–441, 2014.
- [Re16] Reutermann, Jeanine: Differences and Commonalities – A comparative report of video styles and course descriptions on edX, Coursera, Futurelearn and Iversity. In (Khalil, M.; Ebner, M.; Kopp, M.; Lorenz, A., Kalz, M., Hrsg.): *Proceedings of the European MOOC Stakeholder Summit 2016*, Norderstedt: BoD - Books on Demand, S. 383–392, 2016.
- [Se15] Seufert, Sabine: Design Research für die Implementation von eLearning: ein vielversprechendes Paradigma für die Zusammenarbeit von Wissenschaft und Praxis? *HMD Praxis der Wirtschaftsinformatik* 52/1, S. 120–131, 2015.

Qualitätsgesicherte E-Assessments im Student Life Cycle

Xenia V. Jeremias¹, Christian Rabe² und Birgit Sellmer³

Abstract: Das Poster zeigt den Einsatz von E-Assessments entlang des *Student Life Cycle*, auch um auf die heterogenen Kenntnisse der Studierenden zu reagieren. Zunächst werden vielfältige Einsatzszenarien vorgestellt. Im zweiten Teil des Posters werden Aspekte der Qualitätssicherung bei E-Assessments, wie die technische Absicherung, die didaktische Umsetzung der Fragen sowie die Einbindung in Lehre und die Hochschule beleuchtet. Dabei wird auch der E-Assessment-Leitfaden der TH Wildau vorgestellt.

Keywords: diagnostische, formative, summative E-Assessment, Heterogenität, Einsatzszenarien, Qualitätssicherung, strukturelle Einbindung

1 Einleitung

Die TH Wildau bietet als größte Fachhochschule Brandenburgs Studiengänge in Ingenieur- und Naturwissenschaften, Informatik, Wirtschaft und Verwaltung. Die Voraussetzungen, mit denen ein Studium begonnen wird, sind immer breiter gestreut. Um den Bedürfnissen dieser heterogenen Studierendenschaft Rechnung zu tragen, können E-Assessments in vielfältiger Weise – vor und im Studium – eingesetzt werden.

2 Einsatzszenarien

Der Einsatz von E-Assessments erfolgt an der TH Wildau⁴ entlang des *Student Life Cycle*. Dies beginnt in Mathematik vor dem Studium mit Diagnosetests für Studieninteressierte, wobei im Feedback ggf. ein Vorkurs empfohlen wird. Zum Vorkurs Angemeldete werden mit einem Einstufungstest für Tutorien eingeteilt. Dort sowie im TH College werden Online-Selbsttests im Rahmen eines Blended-Learning-Ansatzes genutzt. Damit kann dem individuellen Lerntempo Rechnung getragen werden.

Im Studium werden E-Assessments diagnostisch, formativ und summativ eingesetzt [Mi15]: Zu Studienbeginn können ggf. Lücken diagnostiziert werden, so dass frühzeitig gegengesteuert werden kann. Lehrende können, z. B. als Klausurvorbereitung, Zwischentests einbinden, um Studierende dazu anzuregen, kontinuierlich zu lernen. Im Sinne des *Constructive Alignment* [Bi03] sollte dann auch elektronisch geprüft werden.

¹ TH Wildau, Zentrum f. Qualitätsentwicklung, Hochschulring 1, 15745 Wildau, xenia.jeremias@th-wildau.de

² TH Wildau, Zentrum f. Qualitätsentwicklung, Hochschulring 1, 15745 Wildau, christian.rabe@th-wildau.de

³ TH Wildau, Zentrum f. Qualitätsentwicklung, Hochschulring 1, 15745 Wildau, birgit.sellmer@th-wildau.de

⁴ Stand Sommersemester 2017: 21 Lehrende und 7 Einrichtungen nutzen E-Assessments.

Das Feedback spielt v. a. bei diagnostischen und formativen Tests eine Rolle. Je besser es auf die Lernsituation angepasst ist, desto eher können Studierende damit arbeiten [HS12]. Das Feedback sollte sich nicht auf eine Richtig-Falsch-Rückmeldung begrenzen, denn ohne Erklärung hilft dies meist nicht weiter.

3 Qualitätssicherung von E-Assessments

E-Assessments können nur mit einer umfassenden Qualitätssicherung erfolgreich in die Lehre integriert werden. Zunächst muss die technische Funktion gesichert sein – bei Prüfungen bis zur Verhinderung von Betrug. Wie bei Papierform gilt, dass die Qualität von Assessments durch Beachtung der testdidaktischen Anforderungen erreicht wird. Diese sind in dem Leitfaden www.th-wildau.de/eassessment zusammengefasst.

Bezogen auf die Formen von E-Assessments stellen sich verschiedene Fragen der Qualitätssicherung: Bei diagnostischen Tests gilt es, den Schwierigkeitsgrad auszubalancieren. Die Mindestanforderungen sollten sich in einem solchen Test spiegeln; zur Milderung der Abschreckung jedoch Unterstützung angeboten werden. Formative Tests entfalten ihre Wirkung dann, wenn sie in die Präsenzlehre eingebunden sind. Auf Ergebnisse von Selbsttests sollten Lehrende eingehen und zeigen, dass sie Schlüsse für ihr Modul daraus ableiten. Bei qualitätsgesicherten Prüfungen stimmen die Anforderungen mit den Lernergebnissen aus der Modulbeschreibung überein. Abhängig vom Einsatzszenario können weitere Qualitätskriterien hinzukommen.

Ein weiterer Aspekt der Qualitätssicherung ist die strukturelle Einbindung [Fr15]. So wurde der E-Prüfungsprozess in das Qualitätshandbuch der TH Wildau integriert. Es wurden Vereinbarungen mit den Prüfungskoordinatorinnen der Fachbereiche getroffen und eine passende Online-Anmeldung eingerichtet führt. Des Weiteren stehen Informationen für Studierende und eine E-Prüfung zum Ausprobieren zur Verfügung.

Im Projekt *SOS TEAM – Strukturierung und Optimierung des Selbststudiums. Tutoring, E-Assessment, Mathematik* wird dies im Rahmen des Qualitätspakts Lehre (BMBF, Förderkennzeichen 01PL16042) bearbeitet.

Literaturverzeichnis

- [Bi03] Biggs, J.: *Aligning Teaching and Assessment to Curriculum Objectives*, 2003.
- [Fr15] Friedrich, J.-D. et al.: *E-Assessment als Herausforderung*, S. 8-9, 2015. <https://hochschulforumdigitalisierung.de/veroeffentlichungen>, Stand: 31.05.2017
- [HS12] Handke, J.; Schäfer, A. M.: *E-Learning, E-Teaching und E-Assessment in der Hochschullehre*, S. 43, 150-151, 2012.
- [Mi15] Michel, L.: *Digitales Prüfen und Bewerten im Hochschulbereich*, S. 12, 2015. <https://hochschulforumdigitalisierung.de/veroeffentlichungen>, Stand: 31.05.2017.

Einführung in das wissenschaftliche Arbeiten und Publizieren mittels Blended Learning in der informatischen Fachdidaktik

Peter de Lange, Mohsen Shahriari und Ralf Klamma¹

Abstract: Die Digitalisierung hat auch das wissenschaftliche Arbeiten und Publizieren schon längst erfasst und viele etablierte Praxen werden dadurch transformiert. In der Informatik-Lehre werden hier erste Schritte im Bachelor-Studium, zumeist in Form von Proseminaren, vermittelt. Allerdings unterscheiden sich die Werkzeuge und Vorgehensweise stark von denen, die zeitgleich in den Softwarepraktika vermittelt werden. Wir sind der Überzeugung, dass hier voneinander gelernt werden kann und Studenten besser auf zukünftige Aufgaben im Berufsleben und der Forschung vorbereitet werden können, insbesondere in interdisziplinären Studiengängen wie der Technik-Kommunikation. Wir haben deshalb ein Blended Learning Konzept für ein Proseminar entwickelt, das wir durchführen wollen. Der Kern ist die Erstellung eines quelloffenen Buches mittels Werkzeugen, die für die Erzeugung und Dokumentation von Software verwendet werden, wobei auch Aspekte der Automatisierung von Herstellung und Nutzung berücksichtigt werden. Wir stellen das Konzept und die Elemente vor, auch um Feedback für die Realisierung zu erhalten.

Keywords: Blended Learning, Kollaborative Texterstellung, Open Content

1 Einführung und Motivation

In diesem Beitrag vergleichen wir die Kollaboration im Bereich der Softwareentwicklung und wissenschaftlichen Texterstellung miteinander. Wir sind davon überzeugt, dass Methoden und Techniken der Softwareentwicklungskollaboration [La16] auch im Bereich des Erstellens und Schaffens von wissenschaftlichen Inhalten anwendbar sind, und dieser davon profitiert. Auf der anderen Seite können Softwareentwicklungsprozesse durch Methoden der wissenschaftlichen Dokumentation und kollaborative Textstellungsprozesse verbessert werden. Um dies zu evaluieren beschreiben wir den Entwurf eines Proseminars zur kollaborativen Entwicklung eines quelloffenen Online-Buches, welches wir erstmalig im Sommersemester 2017 durchführen.

2 Aufbau des Kurses und Geplante Evaluierung

Über einen Zeitraum von 12 Wochen werden 25 Kursteilnehmer gemeinsam Inhalte für

¹ RWTH Aachen University, Chair for Information Systems and Databases, Ahornstr. 55, 52074 Aachen, Germany, {lastname}@dbis.rwth-aachen.de

ein Open-Content Online-Buch erstellen und diese gegenseitig begutachten. Als Grundlage und Inspiration für unser Konzept gilt uns die Arbeit von [PK16]. Für das zu entwickelnde Buch nutzen wir GitLab, in Kombination mit dem Static Site Generator Jekyll, welcher den erstellten Inhalt des Git Repositories zu einer Website rendert. Um dem Grundgedanken des Open Content Rechnung zu tragen, sind alle in diesem Kurs genutzten Programme und Techniken Open Source lizenziert. Das Proseminar ist, entsprechend der Theorie des Blended Learning, zweigeteilt. Ein Teil besteht aus Präsenzveranstaltungen, in denen wir semesterbegleitend Einführungen in verschiedene Bereiche des wissenschaftlichen Arbeitens und der kollaborativen Softwareentwicklung geben. Die Veranstaltungen finden für alle Kursteilnehmer gemeinsam statt. Zusätzlich wird jeder Kursteilnehmer je zwei Teams zugewiesen, einem Team welches Inhalte produziert und einem Team welches sich um redaktionelle Arbeiten kümmert. Die Inhalt-produzierenden Teams mit vier Teilnehmern bearbeiten einen Algorithmus aus dem Bereich „Overlapping Community Detection“. Hierzu gibt es verschiedene Bearbeitungsschritte mit Deadlines, zu der der Bearbeitungsstand „online“ sein muss. Auch Peer Reviews sind Teil der Arbeit, wobei jeweils ein Review zu den Resultaten einer anderen Gruppe angefertigt, in einer Präsenzveranstaltung präsentiert und diskutiert wird. Das zweite Team, dem jeder Kursteilnehmer zugewiesen wird, kümmert sich um einen speziellen Aspekt der redaktionellen Arbeit. In unterschiedliche Teams wird sich ein Teil der Teilnehmer mit verschiedenen Aspekten des Buchlayouts beschäftigen. Weitere Teams werden sich um die Animation und Visualisierung der bearbeiteten Algorithmen mittels eines von uns entwickelten Tools [SKK15] kümmern. Ein Team hat zur Aufgabe, den Review Prozess zu organisieren, ein weiteres Team kümmert sich um die Erstellung von Präsentationstemplates. Andere Aufgaben umfassen die Dokumentation des Prozesses und PR-Arbeiten. Für die Koordination der Teams setzen wir Slack ein. Wir wollen evaluieren, inwieweit der kollaborative Ansatz dazu führt, dass Studenten voneinander lernen und sich hierdurch die allgemeine Qualität aller Beiträge verbessert. Dabei werden wir neben qualitativen Daten auch die Daten auch den genannten Systemen nutzen. In einem nächsten Schritt planen wir die Nutzung und Evaluierung von kollaborativen Online-Schreibwerkzeugen. Insgesamt soll die Fachdidaktik des wissenschaftlichen Schreibens in der Informatik mehr Bezug zu aktueller Werkzeugnutzung in der Softwareentwicklung nehmen.

Literaturverzeichnis

- [La16] Lange, de P.; Nicolaescu, P.; Klamma, R.; Koren, I.: DevOpsUse for Rapid Training of Agile Practices within Undergraduate and Startup Communities. In: EC-TEL'16, 2016.
- [PK16] Palaigeorgiou, G.; Kazanidis, I.: Wikis as a Mediation Platform for Developing Learning Communities: The WEKI Framework. In: 1st International Symposium on Emerging Technologies for Education, 2016.
- [SKK15] Shahriari, M.; Krott, S.; Klamma, R.: Disassortative Degree Mixing and Information Diffusion for Overlapping Community Detection in Social Networks (DMID). In: WWW'15 Companion, 2015.

Ein Chatbot für die Erstellung von Übungsaufgaben im E-Learning

Majd Edriss¹, Truong-Sinh An¹, Francois Dubois¹, Agathe Merceron¹

Abstract: Ziel dieser Arbeit ist die Entwicklung eines alternativen Verfahrens zur Erstellung von Übungsaufgaben im Projekt “Smart Learning im Handwerk - SLHw” in Form eines Chatbots. Dieser Beitrag bietet hierzu erste grundlegende Recherche und Lösungsansätze zur Implementierung.

Keywords: Editoren, Question-Test-Interoperability (QTI), Chatbot.

1 Einleitung und verwandte Arbeiten

Im Rahmen des Projekts „Smart Learning im Handwerk - SLHw“ [An16] wurden Editoren zum Erstellen und Pflegen unterschiedlicher Lerninhalte realisiert. Den Medienerstellern ist es möglich mit den Editoren z. B. verschiedene Aufgabentypen zu erstellen und diese in einer Datenhaltung nach der Spezifikation IMS-QTI abzulegen. Dabei wird die Aufgabenstellung mit den Antworten, die Metadaten nach der LOM-Spezifikation und weitere Informationen, wie z. B. die Bewertungen oder die zufällige Anordnung der Antworten, über ein Eingabeformular erfasst. Das Eingabeformular verfolgt den Ansatz aus der Gestaltung von Informationsarchitekturen mit dem Fokus auf die logische Gruppierung, um den umfangreichen Datenmodellen der Lerninhalte mit klarer Bedienbarkeit zu begegnen. Neue Editoren erfordern in der Praxis einen hohen Grad an Einarbeitung, wie der QTI-Editor aus dem SLHw-Projekt gegenüber der Nutzung von MS Word. In diesem Beitrag wird ein neuartiger Ansatz bei der Interaktion mit Editoren in der Form eines *Chatbots* vorgestellt.

Ein Chatbot ist ein textbasiertes Dialogsystem, das mit dem Benutzer in einen Dialog tritt [Di17]. Im hier vorgestellten Ansatz ist ein Chatbot ein regelbasiertes System, um eine von Menschen gestellte Zielsetzung zu realisieren. Dieses Konzept der regelbasierten Konversation greift auch der nicht frei verfügbare Webseiten-Builder Heek [He17] auf. Bei der Erstellung von Webseiten unterstützt Heek den Nutzer, indem ein Dialog auf Frage-Antwort-Basis initiiert wird, bei dem *Etappen* zum Ziel gemeinsam abgearbeitet werden.

¹ Beuth Hochschule für Technik Berlin, Luxemburger Straße 10, 13353 Berlin, s42546@beuth-hochschule.de, truong-sinh.an@beuth-hochschule.de, fdubois@beuth-hochschule.de, merceron@beuth-hochschule.de

2 Ansatz

Der Prototyp des Chatbots soll mit dem Medienersteller textuell in natürlicher Sprache kommunizieren, mit dem Ziel der Erstellung einer Übungsaufgabe. Die Grundidee des Chatbots stellt das sogenannte “Drehbuch” dar. Ein “Drehbuch” bestimmt die Etappen, die der Chatbot durchgeht, um am Ende der Konversation eine Zielsetzung zu erfüllen. Das Diagramm (Abb. 1) zeigt exemplarisch einen Teil des Drehbuchs für die Erstellung einer QTI-Aufgabe. Dieser Prototyp bietet eine Integrationsmöglichkeit für das SLHw Projekt. Hierbei erlaubt der Chatbot die Bearbeitung des Datenmodells, welches den Editoren zugrunde liegt; sämtliche Eingaben des Medienerstellers werden dem entsprechenden Formularfeld zugeordnet. Wenn alle Etappen (Schritte) des Drehbuchs erfolgreich abgeschlossen sind, wird das Ergebnis der Zielsetzung - in diesem Fall die Erstellung einer QTI-Aufgabe - abgespeichert.

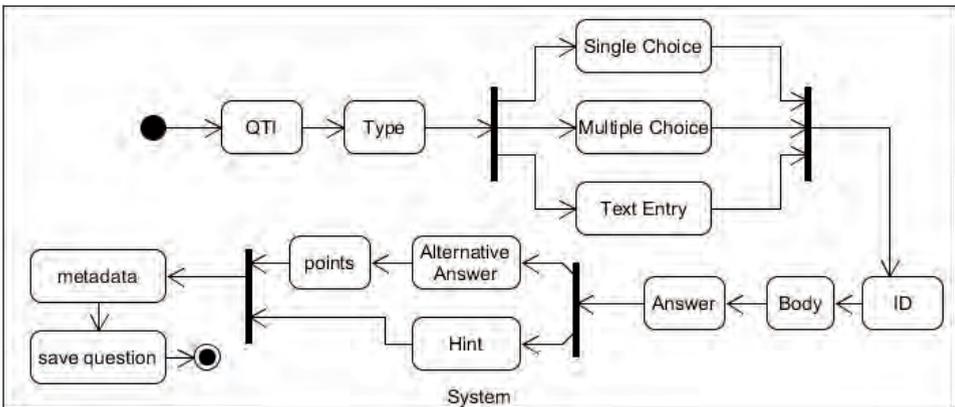


Abb. 1: Szenario für die Erstellung einer QTI-Aufgabe

Danksagung Dieses Vorhaben wird zum Teil mit Mitteln des BMBF Förderkennzeichen 01PD14002B gefördert.

Literaturverzeichnis

- [An16] An, T.-S.; Dubois, F.; Manthey, E.; Merceron, A.: Digitale Infrastruktur und Learning Analytics in Co-Design. In Proceedings of the Workshop Learning Analytics, co-located with the 13th e-Learning Conference of the German Society for Computer Science, Potsdam, Germany, September 11, <http://ceur-ws.org/Vol-1669/>, 2016.
- [Di17] DigitalWiki, <https://www.digitalwiki.de/chatbots>. Aufgerufen: 19.03.2017.
- [He17] Heek SAS, <https://www.heek.com>. Aufgerufen: 21.03.2017.

arsnova.click – App für Live-Quizze im MINT-Unterricht

Christopher Mark Fullarton¹, Tjark Wilhelm Hoeck², Klaus Quibeldey-Cirkel³

Abstract: Die Lehrerfahrungen mit dem Audience-Response-System *ARSnova* (arsnova.eu) in invertierten Vorlesungen führten zu einer Neuausrichtung des Projekts mit Gamification-Elementen [Qu16]. Die Inspiration für die Neuentwicklung *arsnova.click* geht auf *Kahoot!* zurück, eine gamifizierte Quiz-App, die an amerikanischen Schulen verbreitet ist [FW16]. Bei der Realisierung von arsnova.click wurde konsequent ein datenschutzkonformer Weg beschritten: Es werden keine personenbezogenen Daten auf dem Server gehalten. Die Quizfragen mit allen eingebetteten Quellen wie Videos und Bilder verbleiben im Browser der Lehrkraft. Diese innovative Implementierung mittels der Web-Storage-Browsertechnologie garantiert den Datenschutz beim Einsatz an Schulen und Hochschulen in der Europäischen Union. ARSnova ist ein Open-Source-Projekt der THM.

Keywords: Audience Response System, Gamification, Kahoot!, arsnova.click, Datenschutz

1 arsnova.click vs. Kahoot!

arsnova.click steht allen Bildungseinrichtungen zur freien Verfügung, kostenlos und ohne Werbung. Die Nutzung der App erfolgt in beiden User-Rollen – Lehrkraft und Teilnehmer/in – anonym auf derselben Webseite: <https://arsnova.click>. Eine Registrierung ist nicht erforderlich. Auch entfällt eine Installation, da die Quiz-App im BYOD-Browser läuft, designt mobile-first. Die Open-Source-Software⁴ wird von Studierenden der Informatik an der THM entwickelt. Sie steht unter der GNU GPL, kann also auf dem Webserver der Bildungseinrichtung betrieben und für eigene Zwecke angepasst werden.

Im Vergleich zu Kahoot! bietet arsnova.click nicht nur Multiple-Choice und die Umfrage als Frageformate an, sondern auch die numerische Schätzfrage und Kurzantwort. Bis zu 26 Antwortoptionen sind möglich, so lang wie das Alphabet. Es können Emojis, Bilder, Videos und TeX-Formeln in den Fragetext und in die Antwortoptionen eingebunden werden. Zusätzlich gibt es eine Quelltext-Syntaxhervorhebung. Damit eignet sich die App besonders für MINT-Fächer. Quizfrage und Antwortoptionen werden auch auf den Smartphones der Teilnehmenden angezeigt. Beim Abstimmen kann auf einer Prozentkala die *subjektive Antwortsicherheit* mitgeteilt werden: von 0 % „geraten“ bis 100 % „absolut sicher“. Der Mittelwert wird im Ergebnisdigramm angezeigt. arsnova.click ist in Microsoft PowerPoint als Add-in integrierbar. Das Abstimmungsergebnis wird auf der Folie gespeichert und kann als Kurs-Vergleichswert im nächsten Semester dienen.

¹ THM, Fachbereich MNI, Wiesenstraße 14, 35390 Gießen, christopher.mark.fullarton@mni.thm.de

² THM, Fachbereich MNI, Wiesenstraße 14, 35390 Gießen, tjark.wilhelm.hoeck@mni.thm.de

³ THM, Fachbereich MNI, Wiesenstraße 14, 35390 Gießen, klaus.quibeldey-cirkel@mni.thm.de

⁴ Das Open-Source-Projekt wird auf GitHub gehostet: <https://github.com/thm-projects/arsnova.click>.

2 Mediendidaktisches Konzept und Bonusvergabe

Das mediendidaktische Konzept der Quiz-App ist der Einsatz in der Präsenzphase des Inverted-Classroom-Modells (ICM). Als Umfragetool kann es zur Abfrage von Meinungen und Vorwissen eingesetzt werden, als Wissensquiz zur Aktivierung der Studierenden und zur Überprüfung des Verständnisses der Lerninhalte aus der Vorbereitungsphase. Im ICM-Rahmen und nach der Lehrmethode Peer Instruction von Eric Mazur ermöglicht arsnova.click ein *formatives Assessment*: Abhängig vom Ausgang eines Quiz kann die Lehrkraft Lerninhalte wiederholen oder die Vorlesung inhaltlich anpassen. [Qu16]

Die didaktische Besonderheit von arsnova.click ist die Option der *Bonusvergabe*. Extrinsische Anreize in Form von Bonuspunkten, anrechenbar auf die Abschlussprüfung eines Faches, sind weit verbreitet an Hochschulen. Das ARSnova-Team geht hier einen Schritt weiter: Die Bonusvergabe erfolgt im Live-Wettbewerb am Anfang einer Vorlesung, um die Vorbereitung zu prüfen, und am Ende der Vorlesung, um die Aufmerksamkeit und Motivation der Studierenden hochzuhalten. Ein Wissensquiz mit Bonusvergabe am Anfang und Ende einer Vorlesung führt zu mehr und über das Semester anhaltende Präsenz und fördert das aktive Lernen im Hörsaal. Um die Anonymität bei der Projektion der Rangliste zu gewährleisten (durch selbst gewählte oder in Kategorien vorgegebene Nicknamen) und gleichzeitig die Realnamen der Teilnehmenden der Lehrkraft mitzuteilen, bietet arsnova.click eine Authentifizierung an. Dies geschieht über den Authentifizierungsdienst (LDAP) der Bildungseinrichtung. Die Rangliste kann für die Bonusverwaltung als druckfertige Microsoft-Excel-Tabelle exportiert werden.

3 Gamification

Da die Akzeptanz eines Audience Response Systems bei übermäßigem Einsatz und ohne didaktische Einbindung in Lehrmethoden wie ICM und Peer Instruction schwindet, hat das ARSnova-Team multimediale Gamification-Elemente in die App integriert: eine virtuelle *Quiz Lobby*, wo die Nicknamen der eintreffenden Player aufleuchten und alle gemeinsam auf die Quiz-Eröffnung warten, auswählbare Hintergrundmusik in der Lobby und während der Abstimmung, Countdown mit Live-Statistik der abgegebenen Stimmen, animiertes Runterzählen der letzten Sekunden an den Fingern einer Hand und ein lauter Schlusspfeif oder chinesischer Gong. Das Farbschema der App ist frei wählbar.

Literaturverzeichnis

- [FW16] Fallmann, I.; Wala, T.: Die Quizshow im Hörsaal. Studierendenorientiertes Lernen mit Kahoot!. FFH Forschungsforum, Wien, 2016.
- [Qu16] Quibeldey-Cirkel, K.: Lernwiderstände sichtbar machen mit dem Audience Response System ARSnova. In (Knaus T.; Engel, O., Hrsg.): Wi(e)derstände: Digitaler Wandel in Bildungseinrichtungen. Bd. 5, kopaed-Verlag, München, S. 183-198, 2016.

Learning Analytics in Academic Writing Instruction – Opportunities Provided by Thesis Writer (TW)

Christian Rapp, Jakob Ott¹

Abstract: No thesis - no graduation. Academic writing poses manifold challenges to students, instructors and institutions. Consequently, the question arose to what extent academic writing instruction can be supported electronically. With Thesis Writer (TW), a domain-specific, technology-supported learning environment for scaffolding academic writing, combined with an online editor optimised for producing academic text was implemented. Recently, a tracking function was also implemented that allows for fine-grained studying of user interaction with the system, and therefore also their writing and learning processes. In the demonstration, the researchers will present the data types collected, the tracking technology, and the learning analytics opportunities that TW provides.

Keywords: learning analytics, academic writing instruction, instructional technology

Introduction

Thesis Writer (www.thesiswriter.eu) is a domain-specific, technology-supported learning environment for scaffolding academic writing, combined with an online editor optimised for producing academic text [Ra15]. It allows for the scaling of academic writing instruction [RK16]. Following [GP07] TW supports students in particular with: (1) orientation, planning and focusing; (2) proposal writing; (3) text production by tutorials and new linguistic tools including a phrasebank, examples, and linguistic support from an integrated, domain-specific corpus analysis tool; and (4) collaboration and coordination. TW offers functionality that is different to existing automated scoring and feedback tools [AJM15], and is designed for a different genre, specifically the IMRD (Introduction, Methods, Results, and Discussion) scheme [Sw04] which is a standard structure for many theses and research articles.

Recently, a tracking function has been implemented within TW: All user (text) entries and system interactions are logged and timestamped. This allows for the research of support functions provided by TW, to analyse how they are used, and even more interesting, what happens afterwards. For example, it is possible to study if users typed text after using an aid provided or, if they sought other help instead. A currently implemented ‘replay-function’, based on the timestamped data logs (user-system interaction, including text produced in the editor), allows for displaying as if recorded by a screen recorder; however, circumventing problems when utilising screen recording as

¹ Zurich University of Applied Sciences, Center for Innovative Teaching and Learning, School of Management and Law, St., Georgenplatz 2, 8400 Winterthur, Switzerland, rapp@zhaw.ch, ottj@zhaw.ch.

reported by [Ta06]. Therefore, TW will, unobtrusively, allow for research of the following aspects: (1) Usage of TW as an Intelligent Tutoring System; (2) Usage and effect of the aids provided. The effect of the aids can be assessed via corpus analysis, e.g. analysis can include to what extent phrases provided by the phrasebank are incorporated in the text; (3) TW allows real-time collaboration, similar to Google Docs, for instance in teamwork assignments. Here it is possible to analyse interaction among users and the system; and (4) Analyse the text production itself, e.g. time spent on certain sections, revisions processes etc. The described data collection functions allow, firstly, for the in-depth study of writing processes of single users and/or groups (data collection can be confined to groups e.g. classes). Secondly, TW allows for comparison among different users (e.g. social science users in comparison to natural science users, or undergraduates compared to postgraduates) and, given that TW is available in both German and English languages, also between L1 and L2 writers.

TW tracking is triggered using modern XHTML technology from the user's browser. As interaction occurs, entered text is auto-saved and logged to the database every time the user pauses between keystrokes for <0.5 seconds. Logging is performed on a session basis (session is kept alive within TW's text editor), chained entries allow for tracking users' interaction path. To date, TW has created some 140,000 log entries, which include login/-out, navigation, document/group management, writing/revising, and accessing contextual aid functionality (tutorials, phrase base, sample text, linguistic support).

References

- [AJM15] Allen, L.; Jacovina, M.; McNamara, D.: Computer-based writing instruction. In (MacArthur, C.A, Graham, S. & Fitzgerald, J., Hrsg.): Handbook of writing research, The Guildford Press, New York, S. 316-329, 2015.
- [GP07] Graham, S.; Perin, D.: A Meta-Analysis of Writing Instruction for Adolescent Students. *Journal of Educational Psychology* 99/3, S. 445-476, 2007.
- [RK16] Rapp, C.; Kruse, O.: Thesis Writer (TW) – Tapping scale effects in academic writing instruction (demonstration). In Proceedings of the Third (2016) ACM Conference on Learning @ Scale. ACM, New York, S. 325-326, 2016.
- [Ra15] Rapp, C.; Kruse, O.; Erlemann, J.; Ott, J.: Thesis Writer – A System for Supporting Academic Writing. In Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW2015 Companion). ACM, New York, S. 57-60, 2015.
- [Sw04] Swales, J.: Research genres: Explorations and applications (Cambridge applied linguistics series). Cambridge University Press, Cambridge, 2004.
- [Ta06] Tang, J.; Liu, S.; Muller, M.; Lin, J.; Drews, C.: Unobtrusive but invasive: using screen recording to collect field data on computer-mediated interaction. In Proceedings of the 20th Anniversary ACM Conference on Computer Supported Cooperative Work. ACM, Banff, S. 479-482, 2006.

AiX-Analytics: Analytics Tool at RWTH Aachen University

Vlatko Lukarov¹, Ulrik Schroeder¹

Abstract: In this paper we present the AiX-Analytics prototype. The tool is a web-based prototype that visualizes usage statistics and analytics of the learning platform's log data. The prototype has three parts: data management, RESTful application engine, and user interface(s). The data management receives, cleans, analyzes and aggregates the learning platform logs, and saves the derived results and analytics. The RESTful application engine is a Web API application that provides data and analytics to the front-end (UI). The user interface provides a set of visualizations, which interactively presents the learning data for different stakeholders in higher education scenarios. The user interface has two main aspects. The first is intended for the university's administration and eLearning coordinators, in order to help in understanding how different faculties and institutes use the learning platform. The second aspect focuses on individual courses, where the teaching staff and students can take a look which modules, when, and how were used during the ongoing semester(s). The presented work is an ongoing research project at the Learning Technologies Research Group at RWTH Aachen University.

Keywords: Learning Analytics, Decision Making, Learning Data

1 Introduction

This paper presents an analytics tool that provides insights about how students use the platform, resources they appreciate the most, whether they use different devices while learning, and provide insight how they learn with the platform. On the other side, the administration can have an overview how the different faculties use the modules of the learning platform, detect general trends, or invest more resources to improve the eLearning experience. The tool's target audience is divided into two main groups: teachers and students, and administration. This distinction is important because these two main groups have different expectations and goals, although the same data is the basis for the visualizations and indicators. The main emphasis of this prototype is to provide easy-to-understand visualizations of learner-data, and compel the audience to use it in their daily activities.

2 Professors and Students

This prototype builds upon a knowledge gained through previous research on different Learning Analytics prototypes developed and provided as pilot projects at RWTH Aachen

¹ Learning Technologies Research Group, RWTH Aachen University, Ahornstraße 55, 52074 Aachen, Germany, lukarov@cil.rwth-aachen.de, schroeder@informatik.rwth-aachen.de

University [Dy14],[Dy12]. The visualizations in the tool are interactive and enable the user to filter out specific parts, select and zoom-in on other parts of the visualizations. The analytics and visualizations help the teacher to understand how the students use the course and the learning resources. Hence, this prototype provides basic analytics and insights to each course, to inspire both teachers and students to reflect upon their activities in the learning process.

3 Data Driven Decision Making

Decision making is usually based on intuition, presumption, and on accumulated experience, without any specific data or analysis[CDO07]. We implemented a prototype which focuses on the entire platform, its modules, different faculties, and different semesters. The visualizations are interactive and enable the user to focus on specific parts of the visualization. The main idea behind this prototype is to provide actionable intelligence and help the administration in understanding how different faculties learn and use the platform, in order to better distribute resources and support both the staff and the students.

4 Conclusion

AiX Analytics is a web-based application that uses learning generated data to provide analytics in every course room on the platform. Furthermore, it uses the same data to provide analytics and actionable intelligence to the administration, and provide support in the decision making processes in regard to the eLearning initiatives and activities at RWTH Aachen University. We are preparing to conduct a large-scale pilot phase of the prototype to assess its usability, utility, usefulness, and acceptance.

Literature

- [CDO07] Campbell, J.P.; DeBlois, P B.; Oblinger, D G: Academic Analytics. In: Educ. Rev. Bd. 42, S. 40–57, 2007.
- [Dy14] Dyckhoff, A.L.: Action Research and Learning Analytics in Higher Education, 2014.
- [Dy12] Dyckhoff, A.L.; Zielke, D.; Bültmann, M.; Chatti, M. A.; Schroeder, U.: Design and Implementation of a Learning Analytics Toolkit for Teachers Design and Implementation of a Learning Analytics Toolkit for Teachers. In: Source J. Educ. Technol. Soc. Educ. Technol. Soc. Bd. 15, Nr. 153, S. 58–76, 2012.

Mediengestaltung und -einbettung in Moodle zur Vorbereitung von Learning Analytics auf Modulebene

Sandro Leuchter

Keywords: learning analytics, learning tools integration (LTI), Moodle

Learning Analytics (LA), also die Erhebung, Verknüpfung und Nutzung von quantitativen Daten im Lernprozess, verspricht eine höhere Individualisierbarkeit der didaktischen Unterstützung und Beratung von Studierenden und ermöglicht damit einen höheren Lernerfolg [Ar16]. *Learning Management Systeme* (LMS, hier: Moodle) können hierbei prinzipiell als Datenquelle für LA verwendet werden. Die gewonnenen Daten können in unterschiedlichen Szenarien verknüpft werden und u.a. auf Mikroebene innerhalb eines Moduls zur individuellen Unterstützung Studierender benutzt werden [Fe16]. Ebenso kann auf einer Makroebene die Evaluation der Wirksamkeit der didaktischen Methoden eines Moduls anhand der so erhobenen Daten durchgeführt werden. LMS werden als Ermöglicher für LA angesehen, da durch sie vielfältige Daten erzeugt werden und sie bei der Bereitstellung und Analyse solcher Daten helfen könnten. Im Fall des verbreiteten LMS Moodle sind allerdings so gut wie keine LA-Funktionalitäten eingebaut. Die Nutzung von Lernmaterialien gibt einen Hinweis auf den ablaufenden Lernprozess. Analog zu den Annahmen bei der Blickbewegungsmessung ist es plausibel die Dauer der Hinwendung zum Medium als Beleg für Informationsaufnahme, -verarbeitung und deren Nutzung zu nehmen [Rö01].

In Hochschulkursen werden als Lernmaterialien oft Textdokumente und Präsentationen verwendet. SCORM konforme Inhalte werden in der Hochschulpraxis selten benutzt. Dadurch enthält ein typisches Medium eine Vielzahl von zu lernenden Konzepten. Für eine Zuordnung von Mediennutzung zu möglicherweise gelernten Konzepten ist das nicht hilfreich. Daher ist die Modellierung des Wissensstandes und des Lernprozesses schlecht machbar. Die vergleichsweise grobe Granularität der typischen Medien macht sie aber aus studentischer Sicht besser nutzbar. Zur praktischen Anwendbarkeit von Medien in LMS gehört auch die Offline-Nutzbarkeit. Je nach technischer Ausstattung und Verwendungsgewohnheiten ist die Online-Nutzung für Studierende umständlicher oder teurer. Allerdings ist dies eine Voraussetzung für die praktikable Beobachtung des Nutzungsverhaltens.

Anhand eines Pflichtmoduls im Hauptstudium von drei Informatik Bachelor-Studiengängen der Fakultät für Informatik der Hochschule Mannheim, das in den vergangenen zwei Semestern auf eine intensive Einbettung in Moodle ausgerichtet wurde, wurde exploriert, welche Eigenschaften Lernmaterialien LA-freundlich machen und wie Medien beschaffen sein müssen, damit die Interaktion der Lernenden mit dem Medium quantifizierbar wird.

Eine leicht einzusetzende Möglichkeit mit hoher Benutzungsfreundlichkeit für Studierende ist die Verwendung von *Google Slides*. Ein Vorteil dieses webbasierten Präsentationssystems ist, dass ein Satz von Folien als ein einheitliches Objekt verwendbar ist, dass aber jede einzelne Folie darin eine im Fragment-Teil unterschiedliche URL hat. Folienwechsel (analog zu Blickbewegungen zwischen Informationsaufnahme ermöglichenden Fixationen) können so verfolgt werden.

Als Zusatzfunktionalität wurde ein „foliengenaues“ Forum implementiert und über die *Learning Tools Interoperability* Schnittstelle von Moodle in den Kurs integriert. Dadurch haben Studierende einen Anreiz Lernmaterialien online zu verwenden. Das ermöglicht das Tracking der Lernmaterialnutzung auf der Mikroebene. Diese Daten erlauben dann die Aktivitäten im Lernprozess einzuordnen (Vor- und Nachbereitung von Präsenzlehrveranstaltungen, Heranziehen bei Laborübungen und zur selbstständigen Bearbeitung von Übungsaufgaben, Klausurvorbereitung) und vor allem zu quantifizieren (wie lange wurde gelernt, in welchem Tempo und welche Seiten und damit Lerninhalte wurden wann betrachtet).

Im Moment sind die LA-bezogenen Aktivitäten im Modul noch explorativ und es geht um die Untersuchung der technischen Machbarkeit. Im nächsten Schritt sollen die Daten zum einen zur Verbesserung der Unterlagen benutzt werden, indem über alle Kursteilnehmer/innen aggregiert ausgewertet wird, wann im Semester welche Information benutzt wird und ob es dabei Muster gibt. Daneben soll eine individuelle „Heatmap“ generiert werden, die die Konzepte des Kurses räumlich mit einer farblichen Überlagerung der jeweiligen Verweildauern auf den dazu gehörenden Folien anzeigt. Aus didaktischer Sicht wird dabei das Ziel verfolgt, einen Überblick über den gesamten Stoff des Kurses mit einem vermuteten Abgleich des Wissensstandes zu geben. Die sich über die Zeit entwickelnde individuelle Heatmap soll jedem/r Studierenden in regelmäßigen Abständen als Nachricht zugestellt werden, um einen Reflexionsimpuls zu geben.

Zum kommenden Wintersemester ist eine Evaluation geplant. Dabei sollen die LA-Daten mit dem tatsächlichen Lernprozess abgeglichen werden, indem Studierende zusätzlich Lerntagebücher führen, deren Aufzeichnungen dann mit den LA-Daten abgeglichen werden sollen.

Literaturverzeichnis

- [Ar16] Arroway, P. et al.: Learning Analytics in Higher Education. Research report. Louisville, CO: ECAR, March 2016.
- [Fe16] Ferguson, R. et al.: Learning Analytics: Visions of the Future. In: 6th International Learning Analytics and Knowledge (LAK) Conference, 25-29 April 2016, Edinburgh, Scotland, 2016.
- [Rö01] Rötting, M.: Parametersystematik der Augen- und Blickbewegungen für arbeitswissenschaftliche Untersuchungen. Schriftenreihe Rationalisierung und Humanisierung, Bd. 34. Zugl. Dissertation, RWTH Aachen. Aachen: Shaker, 2001.

optes – Mathematik, Lernprozessbegleitung und wie das mit ePortfolios zusammenpasst

Oliver Samoila¹

Abstract: Nach einer kurzen Vorstellung des Verbundprojektes optes „Optimierung der Selbststudiumsphase“ werden folgende Maßnahmen vorgestellt: Zusammenwirken von Test-settings, ePortfolio als Reflexionsinstrument, automatisierte Kompetenzerhebung und Lernprozessbegleitung. Geschlossen wird mit den Möglichkeiten zur Übernahme von entwickelten Materialien, Konzepten und Instrumenten durch andere Bildungseinrichtungen.

Keywords: Mathematik, Lernprozessbegleitung, ePortfolio, Learning Analytics, ILIAS

1 Das Projekt

Übergeordnetes Ziel des Verbundprojektes optes ist, die Studierfähigkeit von StudienanfängerInnen in MINT-Fächern zu erhöhen und ihre mathematischen Kenntnisse auf Studieneingangsniveau zu erweitern, um somit den hohen Abbruchquoten in diesen Studiengängen gezielt entgegenwirken zu können. Fokussiert wird das begleitete Selbststudium in der Studieneingangsphase, d.h. der Zeit vor Studienbeginn, in der sie sich in Vorkursen auf das Studium vorbereiten und das erste Studienjahr, in dem Maßnahmen parallel zu Mathematik-Grundlagenveranstaltungen angeboten werden.

Zum optes-Verbund gehören die Partner Duale Hochschule Baden-Württemberg, Hochschule Ostwestfalen-Lippe, Universität Hamburg, Universität Würzburg und ILIAS open source e-Learning e. V., welche für oben genannte Ziele geeignete Methoden, Konzepte und Werkzeuge entwickeln.

2 Maßnahmen

Die in optes entwickelten Maßnahmen sind alle auf die Unterstützung von Lernenden ausgerichtet oder wirken mittelbar über Lehrende auf sie zurück.

In den Teilprojekten Adaptive mathematische Qualifizierung, ePortfolio, Formatives eAssessment und Propädeutika, eMentoring, eAssessment im Studium und eTutoring werden die Maßnahmen entwickelt. Unterstützt werden diese durch allgemeindidaktische und fachdidaktische Beratung, sowie durch Qualitätssicherung. Die Teilprojekte

¹ Hochschule Ostwestfalen-Lippe, Institut für Kompetenzentwicklung, Liebigstraße 87, 32657 Lemgo, oliver.samoila@hs-owl.de

Dissemination, Koordination Softwareentwicklung, sowie Projekt- und Wissensmanagement vervollständigen das Projekt.

Für die Demonstration im Zuge der DeLFI werden im Folgenden nur die Maßnahmen aus optes beschrieben, die auch live im LMS ILIAS demonstriert werden können. Ein Austausch zu weiteren Themen des optes-Verbundes ist selbstverständlich möglich.

Diagnostische Tests / mathematische Selbsttests: Demonstriert wird ein Auszug aus Tests zu Diagnose- und Trainingszwecken. Neben Multiple-Choice-Aufgaben kommen auch interaktive Aufgabenformate, sowie Aufgaben, die über ein Computer-Algebra-System betrieben werden, zum Einsatz. Ergänzend wird das Gesamtsetting mit Selbstlernmaterialien als Vorkurscurriculum dargestellt.

ePortfolio und Fähigkeitsmatrix: Lernende werden angehalten, während der (begleiteten) Selbstlernphase, ein persönliches ePortfolio zu nutzen. Klassisch werden Reflexionsimpulse über Leitfragen genutzt, welche die Lernenden zur Auseinandersetzung mit dem eigenen Lernen anregen. Besonders in optes: Lernende werden durch zwei neue und individuell bedeutsame Instrumente im ePortfolio unterstützt: Die Fähigkeitsmatrix für Mathematik, die kompetenz- und entwicklungsbezogenen Orientierung bieten soll, und die Lernzielübersicht, welche das individuell relevante Curriculum und den darin vorhandenen Leistungsstand abbildet.

eMentoring – Lernprozessbegleitung als Bindeglied: eMentoren und eMentorinnen betreuen Lernende bei der Bearbeitung ihres ePortfolios und der Interpretation der durch ihre Selbsttestergebnisse gefüllten Fähigkeitsmatrix. Darüber hinaus fördern sie den Aufbau überfachlicher Fähigkeiten bei den Studierenden, indem Sie z.B. Lernstrategien zur Erschließung der mathematischen Inhalte aufzeigen. Vor Ort wird dieses Zusammenwirken veranschaulicht.

3 Weiterverwertung von Angeboten

optes veröffentlicht alle Materialien und Dokumentationen unter Creative Commons-Lizenzen, wobei es sich bei der Mehrheit um OER-konforme Auszeichnungen handelt. Die Materialien können und sollen weiterentwickelt werden, sowie anwendungsbezogenen und wissenschaftlichen Austausch fördern. Die von optes entwickelten Instrumente, für das open-source Lernmanagementsystem ILIAS, stehen der Allgemeinheit ebenfalls zum Gebrauch und zur Weiterentwicklung zur Verfügung.

Literaturverzeichnis

Samoila, O.: Dokumentation zur Erhebung mathematischer Fähigkeiten im Kontext von ePortfolio-Arbeit in optes. Theoretisch-didaktische Konzeption und technisch-methodische Umsetzung, 2016. http://www.optes.de/goto.php?target=wiki_wpage_860_4929 Stand: 22.05.2017.

mod_groupformation: Moodle Plugin zur algorithmisch optimierten Lerngruppenbildung

Johannes Konert¹, René Röpke² und Henrik Bellhäuser³

Abstract: Zur Unterstützung kollaborativer Lernprozesse werden Lernende oftmals in Gruppen eingeteilt und sollen gemeinsam Hausübungen bearbeiten, ein Projekt bewältigen oder eine Präsentation vorbereiten. Mit dem Plugin mod_groupformation können Lehrende automatisiert optimierte Lerngruppen in dem Lernmanagementsystem Moodle bilden. Durch Selbsteinschätzung der Lernenden in Form von Antworten in einem Fragebogen werden Merkmalsvektoren gebildet und mit dem GroupAL Algorithmus optimierte Lerngruppen für die Kriterien Vorwissen, Motivation, Teamorientierung, Persönlichkeitsmerkmale, und präferierte Sprache gebildet. Optional können Lehrende Themen zur Auswahl stellen. Lernende erhalten Feedback zu ihren Persönlichkeitseigenschaften im Vergleich zu ihren Gruppenmitgliedern und dem Kurs.

Keywords: Lerngruppenformation, GroupAL, MoodlePeers

mod_groupformation

Die Eigenschaften der Benutzeroberfläche, sowie die Herleitung der drei unterstützen Szenarien Hausaufgabengruppen, Projektgruppen und Referatsgruppen lassen sich der vorhergehenden Publikation entnehmen [Rö16]. Darin ist auch die Zuordnung der Kriterienauswahl zu den Szenarien und die Herleitung gegeben, wann die Kriterien homogen oder heterogen in der Lerngruppe durch den Algorithmus GroupAL optimiert werden sollen. Das Plugin⁴ unterstützt in der aktuellen Version 1.3 die Sprachen Deutsch und Englisch. Weitere Übersetzungen sind leicht möglich

In Abb. 1 (links) sind die Einstellmöglichkeiten für Dozenten zu sehen. Die Szenario-Auswahl gibt einen kurzen textuellen Hinweis, wann sich welches eignet und welche Kriterien dabei abgefragt und gruppiert werden. Bei Referatsgruppen kann der Dozent Themen zur Auswahl stellen, wodurch ausschließlich nach der Präferenz der Studierenden zu den Themen gruppiert wird. Das Plugin berücksichtigt für Gruppenbildung nur Studierende, für die die Aktivität sichtbar ist. Studierende mit fehlenden Antwortdaten werden zufällig in Gruppen eingeteilt oder können ignoriert

¹ Beuth Hochschule für Technik, FB VI Informatik und Medien, Luxemburger Str. 10, 13353 Berlin, johannes.konert@beuth-hochschule.de

² RWTH Aachen University, Informatik 9 (Learning Technologies), Ahornstraße 55, 52074 Aachen, roepke@informatik.rwth-aachen.de

³ Universität Mainz, Psychologie in den Bildungswissenschaften, Binger Str. 14-16, 55099 Mainz, bellhaeuser@uni-mainz.de

⁴ https://moodle.org/plugins/mod_groupformation/ und https://github.com/moodlepeers/moodle-mod_groupformation/releases, zuletzt abgerufen am 24.03.2017

werden. Dies ist nützlich für Kurse, in denen viele Teilnehmende automatisch eingetragen werden aber passiv sind (und in keine Gruppe kommen sollen).

In Abb. 2 (links) ist die Fragebogenansicht für Studierende zu sehen. Diese wird dynamisch generiert und enthält auch die optionalen Themen oder Vorwissensbereiche, die der Dozent angeben kann. Die weiteren standardisierten Fragen werden auf Basis von hinterlegten XML-Daten generiert. Abb. 2 (rechts) zeigt das sofortige Feedback für Studierende (hier nur Kategorie Persönlichkeitsmerkmale, Angaben in Englisch).



Abb. 1: (links) Einstellungen für Dozenten, (rechts) Vorschau optimierter Gruppenbildung mit Editierfunktion vor Übernahme als Moodle-Gruppen

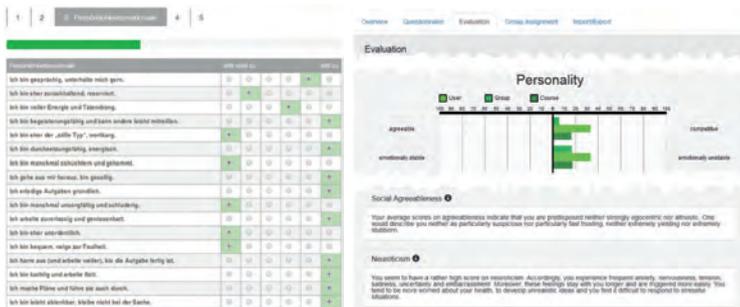


Abb. 2: (links) Dynamisch generierter Fragebogen, (rechts) sofortige Auswertung für Studierende

Literaturverzeichnis

[Rö16] Röpke, R.; Gallwas, E.; Konert, J.; Bellhäuser, H.: MoodlePeers: Automatisierte Lerngruppenbildung auf Grundlage psychologischer Merkmalsausprägungen in E-Learning-Systemen. In: Proc. der 14. E-Learning Fachtagung Informatik der g.i. (DeLFI 2016), S. 233–244, 2016.

Das Web Based Training-Toolkit (WBT-T)

David Weiß¹, Patrick Sacher², Daniel Schiffner³

Abstract: Das WBT-Toolkit ist ein Framework, mit dem man beliebige Webseiten mit WBT-Funktionalitäten erweitern kann. Dadurch ist es möglich, sich bei der Entwicklung von WBTs komplett auf das Frontend und den Entwurf des Lerninhaltes zu konzentrieren. WBT-spezifische Backend-Funktionalitäten, wie z.B. SCORM- und xAPI-Kompatibilität, Fragenlogik und Speicherung des Lernfortschrittes, werden automatisch integriert.

Keywords: Web Based Training, WBT, Autorensysteme, JavaScript-Toolkit, LernBar

1 Einleitung

Zur Produktion von Lerneinheiten (WBTs), stehen viele dedizierte Autorensysteme zur Verfügung, die Autoren bei der Strukturierung und Erstellung von Lerninhalten unterstützen. Eines dieser Autorensysteme ist die, an der Goethe-Universität seit 2004 entwickelte, LernBar [Da17].

Inhalte sind jedoch oft bereits vorhanden und können wiederverwendet werden (siehe z.B.: Creative Commons (CC) [Cr15] oder Online Educational Resources (OER)). Zusätzlich steht den meisten Autoren bereits ein Content Management System (CMS) zur Verfügung, dessen Workflow bekannt ist und viele Features bezüglich der Medienproduktion bietet. eLearning-spezifische Funktionen fehlen jedoch meist.

Gemeinsam mit dem Startup *liquidmoon* entwickelten wir deshalb ein WBT-Toolkit, welches es ermöglicht eLearning-spezifische Funktionen in beliebige Webinhalte zu integrieren. Mit Hilfe dieses Toolkits entstanden mehrere umfangreiche und interaktive Lerneinheiten für eine große Bankgesellschaft, deren Frontend (Responsives Design, One-Pager, Slider etc.) komplett losgelöst von dem WBT-spezifischen Backend entwickelt werden konnte. Funktionalitäten, wie die Fragen(-übermittlung), Metadaten [Du12, IE09] sowie xAPI und SCORM-Kompatibilität, wurden durch das WBT-Toolkit nachgereicht.

¹ Goethe-Universität Frankfurt, studiumdigitale, Varrentrappstr. 40-42, 60486 Frankfurt am Main, weiss@studiumdigitale.uni-frankfurt.de

² Goethe-Universität Frankfurt, studiumdigitale, Robert-Mayer-Str. 10, 60325 Frankfurt am Main, sacher@studiumdigitale.uni-frankfurt.de

³ Goethe-Universität Frankfurt, studiumdigitale, Robert-Mayer-Str. 10, 60325 Frankfurt am Main, schiffner@studiumdigitale.uni-frankfurt.de

2 Das WBT-Toolkit

Bei dem WBT-Toolkit handelt es sich zum einen um eine Low-Level JavaScript-API und zum anderen um eigenen HTML-Bausteinen, mit deren Hilfe einzelne Segmente einer bestehenden Seite funktional erweitert werden können. Durch die Verwendung von CustomElements⁴ wird eine hohe Modularisierung sowie progressive Enhancement erreicht.

Am Beispiel einer SingleChoice Frage ergibt sich folgender Ablauf zur Integration des WBT-Toolkits:

1. Einbinden des WBT-Toolkits im Header der bestehenden Seite.
2. Laden einer (optionalen) Konfiguration oder Convenience-Funktionen.
3. Hinzufügen des CustomElements `<lb-singlechoice>` um betroffene `input[type=radio]` Elemente um Punkte- und Korrektheits-Attribute zu erweitern.

Danach ist es möglich mit `[lb-sc].getState()` oder `[lb-sc].getQuestionScore()` die aktuelle Auswahl oder die aktuell erreichten Punkte anzufragen. Mit der Low-Level API können diese Informationen dann als SCORM oder xAPI an ein LRS übermittelt werden. Eine Erstellung von eigenen xAPI-Statements ist ebenfalls möglich.

Da das WBT-Toolkit unabhängig von dem verwendeten Frontend eingesetzt werden kann, ergeben sich ganz neue Einsatzmöglichkeiten. Neuheiten, welche sich durch die Weiterentwicklung des WWWs ergeben, sind direkt einsetzbar und können mit WBT-Funktionalitäten angereichert werden. Somit ist die Rolle eines klassischen WBT-Autorensystems nicht mehr, sich an ständig ändernder Frontendbedingungen (OnePager, Parallax Effekte, Responsive Design etc.) anzupassen, vielmehr können sich Entwickler mit Hilfe eines WBT-Toolkits mehr auf das konzentrieren, was WBTs im Vergleich zum sonstigen Internet besonders macht: den Lehr-/Lern-Aspekt.

Literaturverzeichnis

- [Cr15] Creative Commons '15, <https://stateof.creativecommons.org/2015/>, Stand: 17.07.2017.
- [Da17] Das LernBar Autorensystem 2017, <http://lernbar.uni-frankfurt.de/>, Stand: 17.07.2017.
- [Du12] Dublin Core Metadata Initiative. Dublin core metadata element set, version 1.1, 2012.
- [IE09] IEEE Standards Association 2009, 1484.12.1-2002 - IEEE Standard for Learning Object Metadata, <https://standards.ieee.org/findstds/standard/1484.12.1-2002.html>, Stand: 17.07.2017.

⁴ <https://www.w3.org/TR/custom-elements/>

Verzeichnis der Autorinnen und Autoren

Alkathib, Wael	15	Goedicke, Michael	75, 261
An, Truong-Sinh.....	345, 387	Grella, Catrina.....	87
Arnold, Uwe.....	303	Hagel, Georg.....	125
Aschmann, Olaf	15	Hagerer, Ilse	143
Bach, Carlo	185	Hanck, Christoph	75
Baedorf, Dominik	377	Hara, Tenshi.....	215
Bauer, Mathias	267	Hartung, Lukas.....	155
Bellhäuser, Henrik	309, 399	Haupt, Thomas.....	303
Bergner, Nadine	117	Hey, Thorsten.....	111
Blasberg, Alexander.....	75	Herczeg, Michael	357
Boos, Maria.....	377	Hielscher, Michael	131
Braun, Iris	215	Hobert, Sebastian	339
Breiter, Andreas	63	Hoeck, Tjark Wilhelm	389
Chatti, Mohamed Amine.....	51	Hoppe, Uwe	143
Dehne, Julian.....	167	Host, Henning	377
de Lang, Peter	385	Igel, Christoph.....	111
Dick, Uwe	267	Jacobsen, Finn.....	357
Dilger, Bernadette	375	Jaster, Svenja.....	377
Döbeli Honegger, Beat	131	Jeremias, Xenia V.....	383
Dorochevsky, Michael	267	Kapp, Felix.....	215
Dubois, Francois	345, 387	Karn, Nils.....	87
Eckhardt, Linda	27	Katzlinger-Felhofer, Elisabeth...	197
Edriss, Majd	285	Kaufmann, Luise.....	369
Ehlenz, Matthias	297	Kämmerer, Frauke	155
Ellek, Dominic	387	Kennedy, Ian G.	333
Finken, Julia.....	63	Kerber, Florian.....	191
Freier, Pascal.....	339	Klamma, Ralf.....	385
Fullarton, Christopher Mark	389	Kohls, Christian	39
Gabler, Doris.....	125	Konert, Johannes.....	309, 399

Koschmider, Agnes.....	99	Münster, Guido	39
König, Stefan	27	Neitzel, Svenja	203
Kramer, Matthias	261	Noichl, Svenja.....	117
Kreggenfeld, Niklas	137	Ott, Jakob	391
Krieter, Philipp.....	63	Ott-Kroner, Alexandra	267
Kubica, Tommy	215	Otto, Benjamin.....	75
Kugler, Christiane	111	Pawlitzek, René.....	185
Kuhlenkötter, Bernd.....	137	Pengel, Norbert	179
Küppers, Bastian	191	Peter, Daniel.....	111
Le, Nguyen-Thanh.....	249	Pilak, Andreas	285
Leonhardt, Thimo.....	297	Pinkwart, Nils	227, 249, 321
Leuchter, Sandro	395	Pollok, Falk	51
Liere-Netheler, Kirsten	143	Prasse, Doreen.....	131
Lohner, David	379, 381	Prinz, Christopher	137
Löhr, Manuel.....	285	Quibeldey-Cirkel, Klaus	389
Lucke, Ulrike	167	Rabe, Christian.....	383
Lukarov, Vlatko	239, 393	Rapp, Christian	375, 391
Martens, Thomas.....	267	Rau, Johannes	285
Marx, Franziska	63	Rebholz, Sandra	27
Massing, Till	75	Reckmann, Natalie.....	75
Mattner, Frauke.....	111	Reichelt, Maria.....	303
Meinel, Christoph	87	Reiners, Almut	339
Meng, Michael	111	Rensing, Christoph.....	15, 203, 309
Merceron, Agathe	345, 387	Renz, Jan	87
Meyer, Michaela	63	Robra-Bissantz, Susanne.....	285
Meyer, Ulrike.....	191	Röpke, René.....	309, 399
Mueller, Rainer	375	Rupp, Simon	185
Müller-Amthor, Martina	125	Sacher, Patrick	401
Müller, Wolfgang.....	27	Samoila, Oliver	397

Schaarschmidt, Mario	99	Wollersheim, Heinz-Werner	179
Schiffner, Daniel	401	Wolters, Christian	357
Schill, Alexander.....	215	Zheng, Zhilin.....	321
Schimkat, Ralf.....	375	Zollenkopf, Max	15
Schneider, Oliver	267		
Schroeder, Ulrik.....	51, 117, 191,239, 297, 393		
Schumann, Sandy.....	75		
Schumann, Matthias	339		
Schwinning, Nils.....	75		
Sellmer, Birgit.....	383		
Shahriari, Mohsen	385		
Slawik, Isabel.....	273		
Söbke, Heinrich.....	155, 303		
Stabauer, Martin.....	197		
Strickroth, Sven.....	227		
Striewe, Michael	261		
Thor, Andreas	179, 369		
Tittel, Stephan.....	15		
Tschudi, Christof.....	131		
Ullrich, Carsten	137		
van Treel, Peer	285		
Vogelsang, Kristin	143		
Vossen, Paul H.....	333		
Wartschinski, Laura	249		
Weiß, David.....	401		
Welz, Tobias	369		
Wessel, Daniel	357		
Wiepke, Axel	167		

GI-Edition Lecture Notes in Informatics

- P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.
- P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.
- P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB'2001.
- P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensorgestützte Chirurgie; Workshop des SFB 414.
- P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.
- P-6 Hans-Jürgen Appelpath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.
- P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.
- P-8 Reinhard Keil-Slawik, Johannes Magenheimer (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.
- P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.
- P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.
- P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.
- P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.
- P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.
- P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.
- P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.
- P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.
- P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.
- P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.
- P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.
- P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungsband).
- P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.
- P-22 Sigrid Schubert, Johannes Magenheimer, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.
- P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit
- P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen
- P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003
- P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web
- P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin
- P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Erfahrungen und Visionen
- P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems
- P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications
- P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures

- P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003
- P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft
- P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)
- P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenber, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)
- P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenber (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik
- P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik
- P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003
- P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003
- P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004
- P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing
- P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste
- P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications
- P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services
- P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004
- P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment
- P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society
- P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications
- P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven
- P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm
- P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik
- P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004
- P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration
- P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Ad-hoc-Netze Medienintegration
- P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government
- P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen
- P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality
- P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Rannenber, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments
- P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung
- P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen
- P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit
- P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and its Applications

- P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005
- P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web
- P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik
- P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)
- P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)
- P-69 Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005
- P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)
- P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005
- P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment
- P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"
- P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology
- P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture
- P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz
- P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006
- P-78 K.-O. Wenkel, P. Wagner, M. Morgens-tern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel
- P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006
- P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce
- P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS'06
- P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006
- P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics
- P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications
- P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems
- P-86 Robert Krimmer (Ed.): Electronic Voting 2006
- P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik
- P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006
- P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur
- P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006
- P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006
- P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1
- P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2
- P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen
- P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies
- P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

- P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices
- P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik
- P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik
- P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten
- P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments
- P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)
- P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme
- P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007
- P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops
- P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications
- P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures
- P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1
- P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2
- P-111 Christian Eibl, Johannes Magenheimer, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik
- P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis
- P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)
- P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics
- P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walthert (Eds.) German conference on bioinformatics GCB 2007
- P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007
- P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and management The 4th International Conference on Grid Service Engineering and Management GSEM 2007
- P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007
- P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications
- P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges
- P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik
- P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

- P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)
- P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Special Interest Groups PARS and PARVA
- P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT: Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung
- P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)
- P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008
- P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany
- P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems
- P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik und E-Voting, CC
- P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik
- P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 1
- P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik Band 2
- P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik – Aktuelle Forschungsergebnisse
- P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008
- P-137 Arslan Brömme, Christoph Busch, Detlef Hühlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures
- P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008
- P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschaftsinformatik
- P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics
- P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management
- P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung
- P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

- P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)
- P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management
- P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)
- P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009
- P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I²CS 2009
- P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum
Kommunikationstechnologien
Beiträge der Fachtagung
- P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband
- P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for
Enterprise Software
- P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop EMISA 2009
- P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning
Fachtagung Informatik
- P-154 Stefan Fischer, Erik Maehle, Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben
- P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures
- P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«
- P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics
2009
- P-158 W. Claudepein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture
Reloaded – Informationsgestützte
Landwirtschaft
- P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010
- P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)
Software Engineering 2010 –
Workshopband
(inkl. Doktorandensymposium)
- P-161 Gregor Engels, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)
Modellierung 2010
- P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung
Verwaltungsinformatik (FTVI) und
Fachtagung Rechtsinformatik (FTRI) 2010
- P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre
Informationssysteme
Technologien, Anwendungen und
Dienste zur Unterstützung von mobiler
Kollaboration
- P-164 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2010: Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

- P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)
10th International Conference on Innovative Internet Community Systems (I²CS) – Jubilee Edition 2010 –
- P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
3. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung
- P-167 Robert Krimmer, Rüdiger Grimm (Eds.)
4th International Conference on Electronic Voting 2010
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik
- P-169 Michael Kerres, Nadine Ojstersek Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.
- P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit
- P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management
- P-172 Stefan Klink, Agnes Koschmider Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)
- P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010
- P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGktive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“
- P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1
- P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2
- P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC
- P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschafts-informatik (WI-MAW), Aachen, 2010
- P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM 2010
Second International Workshop on Formal Methods and Agile Methods
- P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)
- P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft
- P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action
- P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

- P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)
- P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I²CS)
- P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikationstechnologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn
- P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden
- P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf INFOS 2011
14. GI-Fachtagung Informatik und Schule
- P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)
- P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group
- P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities
- P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011
- P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landwirtschaft Fokus Forstwirtschaft
- P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
- P-196 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2012
Proceedings of the 11th International Conference of the Biometrics Special Interest Group
- P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012
- P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband
- P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops
- P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012
- P-202 Andrea Back, Markus Bick, Martin Breunig, Key Poustchi, Frédéric Thiesse (Hrsg.)
MMS 2012: Mobile und Ubiquitäre Informationssysteme
- P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
5. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung
- P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)
12th International Conference on Innovative Internet Community Systems (I²CS 2012)
- P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)
5th International Conference on Electronic Voting 2012 (EVOTE2012)
Co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC
- P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)
EMISA 2012
Der Mensch im Zentrum der Modellierung
- P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.
24.–26. September 2012

- P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.)
INFORMATIK 2012
- P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.)
Nachhaltiges Software Management
- P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.)
Automotive – Safety & Security 2012
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
- P-211 M. Clasen, K. C. Kersebaum, A. Meyer-Aurich, B. Theuvsen (Hrsg.)
Massendatenmanagement in der Agrar- und Ernährungswirtschaft
Erhebung - Verarbeitung - Nutzung
Referate der 33. GIL-Jahrestagung
20. – 21. Februar 2013, Potsdam
- P-212 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2013
Proceedings of the 12th International Conference of the Biometrics Special Interest Group
04.–06. September 2013
Darmstadt, Germany
- P-213 Stefan Kowalewski, Bernhard Rumpe (Hrsg.)
Software Engineering 2013
Fachtagung des GI-Fachbereichs Softwaretechnik
- P-214 Volker Markl, Gunter Saake, Kai-Uwe Sattler, Gregor Hackenbroich, Bernhard Mitschang, Theo Härder, Veit Köppen (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 2013
13. – 15. März 2013, Magdeburg
- P-215 Stefan Wagner, Horst Lichter (Hrsg.)
Software Engineering 2013
Workshopband
(inkl. Doktorandensymposium)
26. Februar – 1. März 2013, Aachen
- P-216 Gunter Saake, Andreas Henrich, Wolfgang Lehner, Thomas Neumann, Veit Köppen (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW) 2013 – Workshopband
11. – 12. März 2013, Magdeburg
- P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
6. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung
03.–04. Juni 2013, Erlangen
- P-218 Andreas Breiter, Christoph Rensing (Hrsg.)
DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
8. – 11. September 2013, Bremen
- P-219 Norbert Breier, Peer Stechert, Thomas Wilke (Hrsg.)
Informatik erweitert Horizonte
INFOS 2013
15. GI-Fachtagung Informatik und Schule
26. – 28. September 2013
- P-220 Matthias Horbach (Hrsg.)
INFORMATIK 2013
Informatik angepasst an Mensch, Organisation und Umwelt
16. – 20. September 2013, Koblenz
- P-221 Maria A. Wimmer, Marijn Janssen, Ann Macintosh, Hans Jochen Scholl, Efthimos Tambouris (Eds.)
Electronic Government and Electronic Participation
Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013
16. – 19. September 2013, Koblenz
- P-222 Reinhard Jung, Manfred Reichert (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2013)
St. Gallen, Switzerland
September 5. – 6. 2013
- P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2013
10. – 11. September 2013
Kloster Banz, Germany
- P-224 Eckhart Hanser, Martin Mikusz, Masud Fazal-Baqaie (Hrsg.)
Vorgehensmodelle 2013
Vorgehensmodelle – Anspruch und Wirklichkeit
20. Tagung der Fachgruppe Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik (WI-VM) der Gesellschaft für Informatik e.V.
Lörrach, 2013
- P-225 Hans-Georg Fill, Dimitris Karagiannis, Ulrich Reimer (Hrsg.)
Modellierung 2014
19. – 21. März 2014, Wien
- P-226 M. Clasen, M. Hamer, S. Lehnert, B. Petersen, B. Theuvsen (Hrsg.)
IT-Standards in der Agrar- und Ernährungswirtschaft Fokus: Risiko- und Krisenmanagement
Referate der 34. GIL-Jahrestagung
24. – 25. Februar 2014, Bonn

- P-227 Wilhelm Hasselbring,
Nils Christian Ehmke (Hrsg.)
Software Engineering 2014
Fachtagung des GI-Fachbereichs
Softwaretechnik
25. – 28. Februar 2014
Kiel, Deutschland
- P-228 Stefan Katzenbeisser, Volkmar Lotz,
Edgar Weippl (Hrsg.)
Sicherheit 2014
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 7. Jahrestagung des
Fachbereichs Sicherheit der
Gesellschaft für Informatik e.V. (GI)
19. – 21. März 2014, Wien
- P-229 Dagmar Lück-Schneider, Thomas
Gordon, Siegfried Kaiser, Jörn von
Lucke, Erich Schweighofer, Maria
A. Wimmer, Martin G. Löhle (Hrsg.)
Gemeinsam Electronic Government
ziel(gruppen)gerecht gestalten und
organisieren
Gemeinsame Fachtagung
Verwaltungsinformatik (FTVI) und
Fachtagung Rechtsinformatik (FTRI)
2014, 20.-21. März 2014 in Berlin
- P-230 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2014
Proceedings of the 13th International
Conference of the Biometrics Special
Interest Group
10. – 12. September 2014 in
Darmstadt, Germany
- P-231 Paul Müller, Bernhard Neumair,
Helmut Reiser, Gabi Dreo Rodosek
(Hrsg.)
7. DFN-Forum
Kommunikationstechnologien
16. – 17. Juni 2014
Fulda
- P-232 E. Plödereder, L. Grunske, E. Schneider,
D. Ull (Hrsg.)
INFORMATIK 2014
Big Data – Komplexität meistern
22. – 26. September 2014
Stuttgart
- P-233 Stephan Trahasch, Rolf Plötzner, Gerhard
Schneider, Claudia Gayer, Daniel Sassiati,
Nicole Wöhrle (Hrsg.)
DeLFI 2014 – Die 12. e-Learning
Fachtagung Informatik
der Gesellschaft für Informatik e.V.
15. – 17. September 2014
Freiburg
- P-234 Fernand Feltz, Bela Mutschler, Benoît
Ottjacques (Eds.)
Enterprise Modelling and Information
Systems Architectures
(EMISA 2014)
Luxembourg, September 25-26, 2014
- P-235 Robert Giegerich,
Ralf Hofestädt,
Tim W. Nattkemper (Eds.)
German Conference on
Bioinformatics 2014
September 28 – October 1
Bielefeld, Germany
- P-236 Martin Engstler, Eckhart Hanser,
Martin Mikusz, Georg Herzwurm (Hrsg.)
Projektmanagement und
Vorgehensmodelle 2014
Soziale Aspekte und Standardisierung
Gemeinsame Tagung der Fachgruppen
Projektmanagement (WI-PM) und
Vorgehensmodelle (WI-VM) im
Fachgebiet Wirtschaftsinformatik der
Gesellschaft für Informatik e.V., Stuttgart
2014
- P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2014
4.–6. November 2014
Stuttgart, Germany
- P-238 Arno Ruckelshausen, Hans-Peter
Schwarz, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und
Ernährungswirtschaft
Referate der 35. GIL-Jahrestagung
23. – 24. Februar 2015, Geisenheim
- P-239 Uwe Aßmann, Birgit Demuth, Thorsten
Spitta, Georg Püschel, Ronny Kaiser
(Hrsg.)
Software Engineering & Management
2015
17.-20. März 2015, Dresden
- P-240 Herbert Klenk, Hubert B. Keller, Erhard
Plödereder, Peter Dencker (Hrsg.)
Automotive – Safety & Security 2015
Sicherheit und Zuverlässigkeit für
automobile Informationstechnik
21.–22. April 2015, Stuttgart
- P-241 Thomas Seidl, Norbert Ritter,
Harald Schöning, Kai-Uwe Sattler,
Theo Härder, Steffen Friedrich,
Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW 2015)
04. – 06. März 2015, Hamburg

- P-242 Norbert Ritter, Andreas Henrich, Wolfgang Lehner, Andreas Thor, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015) – Workshopband
02. – 03. März 2015, Hamburg
- P-243 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
8. DFN-Forum
Kommunikationstechnologien
06.–09. Juni 2015, Lübeck
- P-244 Alfred Zimmermann, Alexander Rossmann (Eds.)
Digital Enterprise Computing (DEC 2015)
Böblingen, Germany June 25-26, 2015
- P-245 Arslan Brömme, Christoph Busch, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2015
Proceedings of the 14th International Conference of the Biometrics Special Interest Group
09.–11. September 2015
Darmstadt, Germany
- P-246 Douglas W. Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.)
INFORMATIK 2015
28.9.-2.10. 2015, Cottbus
- P-247 Hans Pongratz, Reinhard Keil (Hrsg.)
DeLFI 2015 – Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
1.–4. September 2015
München
- P-248 Jens Kolb, Henrik Leopold, Jan Mendling (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 6th Int. Workshop on Enterprise Modelling and Information Systems Architectures, Innsbruck, Austria
September 3-4, 2015
- P-249 Jens Gallenbacher (Hrsg.)
Informatik
allgemeinbildend begreifen
INFOS 2015 16. GI-Fachtagung
Informatik und Schule
20.–23. September 2015
- P-250 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2015
Hybride Projektstrukturen erfolgreich umsetzen
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Elmshorn 2015
- P-251 Detlef Hühnlein, Heiko Roßnagel, Raik Kuhlisch, Jan Ziesing (Eds.)
Open Identity Summit 2015
10.–11. November 2015
Berlin, Germany
- P-252 Jens Knoop, Uwe Zdun (Hrsg.)
Software Engineering 2016
Fachtagung des GI-Fachbereichs Softwaretechnik
23.–26. Februar 2016, Wien
- P-253 A. Ruckelshausen, A. Meyer-Aurich, T. Rath, G. Recke, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Intelligente Systeme – Stand der Technik und neue Möglichkeiten
Referate der 36. GIL-Jahrestagung
22.-23. Februar 2016, Osnabrück
- P-254 Andreas Oberweis, Ralf Reussner (Hrsg.)
Modellierung 2016
2.–4. März 2016, Karlsruhe
- P-255 Stefanie Betz, Ulrich Reimer (Hrsg.)
Modellierung 2016 Workshopband
2.–4. März 2016, Karlsruhe
- P-256 Michael Meier, Delphine Reinhardt, Steffen Wendzel (Hrsg.)
Sicherheit 2016
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 8. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
5.–7. April 2016, Bonn
- P-257 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
9. DFN-Forum
Kommunikationstechnologien
31. Mai – 01. Juni 2016, Rostock

- P-258 Dieter Hertweck, Christian Decker (Eds.)
Digital Enterprise Computing (DEC 2016)
14.–15. Juni 2016, Böblingen
- P-259 Heinrich C. Mayr, Martin Pinzger (Hrsg.)
INFORMATIK 2016
26.–30. September 2016, Klagenfurt
- P-260 Arslan Brömme, Christoph Busch,
Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2016
Proceedings of the 15th International
Conference of the Biometrics Special
Interest Group
21.–23. September 2016, Darmstadt
- P-261 Detlef Rätz, Michael Breidung, Dagmar
Lück-Schneider, Siegfried Kaiser, Erich
Schweighofer (Hrsg.)
Digitale Transformation: Methoden,
Kompetenzen und Technologien für die
Verwaltung
Gemeinsame Fachtagung
Verwaltungsinformatik (FTVI) und
Fachtagung Rechtsinformatik (FTRI) 2016
22.–23. September 2016, Dresden
- P-262 Ulrike Lucke, Andreas Schwill,
Raphael Zender (Hrsg.)
DeLFI 2016 – Die 14. E-Learning
Fachtagung Informatik
der Gesellschaft für Informatik e.V. (GI)
11.–14. September 2016, Potsdam
- P-263 Martin Engstler, Masud Fazal-Baqaie,
Eckhart Hanser, Oliver Linsen, Martin
Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und
Vorgehensmodelle 2016
Arbeiten in hybriden Projekten: Das
Sowohl-als-auch von Stabilität und
Dynamik
Gemeinsame Tagung der Fachgruppen
Projektmanagement (WI-PM) und
Vorgehensmodelle (WI-VM) im
Fachgebiet Wirtschaftsinformatik
der Gesellschaft für Informatik e.V.,
Paderborn 2016
- P-264 Detlef Hühnlein, Heiko Roßnagel,
Christian H. Schunck, Maurizio Talamo
(Eds.)
Open Identity Summit 2016
der Gesellschaft für Informatik e.V. (GI)
13.–14. October 2016, Rome, Italy
- P-265 Bernhard Mitschang, Daniela
Nicklas, Frank Leymann, Harald
Schöning, Melanie Herschel, Jens
Teubner, Theo Härder, Oliver Kopp,
Matthias Wieland (Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW 2017)
6.–10. März 2017, Stuttgart
- P-266 Bernhard Mitschang, Norbert Ritter,
Holger Schwarz, Meike Klettke, Andreas
Thor, Oliver Kopp, Matthias Wieland
(Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW 2017)
Workshopband
6.–7. März 2017, Stuttgart
- P-267 Jan Jürjens, Kurt Schneider (Hrsg.)
Software Engineering 2017
21.–24. Februar 2017, Hannover
- P-268 A. Ruckelshausen, A. Meyer-Aurich,
W. Lentz, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und
Ernährungswirtschaft
Fokus: Digitale Transformation –
Wege in eine zukunftsfähige
Landwirtschaft
Referate der 37. GIL-Jahrestagung
06.–07. März 2017, Dresden
- P-269 Peter Dencker, Herbert Klenk, Hubert
Keller, Erhard Plödereder (Hrsg.)
Automotive – Safety & Security 2017
30.–31. Mai 2017, Stuttgart
- P-271 Paul Müller, Bernhard Neumair, Helmut
Reiser, Gabi Dreö Rodosek (Hrsg.)
10. DFN-Forum Kommunikations-
technologien
30. – 31. Mai 2017, Berlin
- P-272 Alexander Rossmann, Alfred
Zimmermann (eds.)
Digital Enterprise Computing
(DEC 2017)
11.–12. Juli 2017, Böblingen

P-273 Christoph Igel, Carsten Ullrich,
Martin Wessner (Hrsg.)
BILDUNGSRÄUME
DeLFI 2017
Die 15. e-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V. (GI)
5. bis 8. September 2017, Chemnitz

The titles can be purchased at:
Köllen Druck + Verlag GmbH
Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn
Fax: +49 (0)228/9898222
E-Mail: druckverlag@koellen.de

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into

- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: <http://www.gi.de/service/publikationen/lni/>

ISSN 1617-5468

ISBN 978-3-88579-667-1

This volume is the proceedings of the fifteenth e-learning conference of the Gesellschaft für Informatik (GI). The carefully reviewed contributions reflect the state of the art in various areas of e-learning and educational technologies in computer science and related areas, including cooperation and collaboration, content engineering and management, architectures and tools, case studies, organizational and didactical aspects, educational models and environments.