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 978-3-88579-691-6

"DELFI 2019" is the seventeenth event in a conference series focusing on the discussion of the newest issues arising from research and practical experience in the field of technology enhanced learning. The carefully reviewed contributions reflect state of the art in various technological application aspects within computer science and related areas, including content engineering and content management, architecture and tools, cooperation and collaboration, case studies, education models and environments, and educational technology.

N. Pinkwart, J. Konert (Hrsg.): DELFI 2019

297

# GI-Edition

## Lecture Notes in Informatics

Niels Pinkwart, Johannes Konert (Hrsg.)

# DELFI 2019 – Die 17. Fachtagung Bildungstechnologien

16.–19. September 2019
Berlin

Proceedings


GESELLSCHAFT
FÜR INFORMATIK

Niels Pinkwart, Johannes Konert (Hrsg.)

# DELFI 2019 –
# Die 17. Fachtagung Bildungstechnologien

**16. – 19. September 2019**
**Berlin, Deutschland**

Gesellschaft für Informatik e.V. (GI)

**Lecture Notes in Informatics (LNI) - Proceedings**
Series of the Gesellschaft für Informatik (GI)

Volume P-297

ISBN 978-3-88579-691-6
ISSN 1617-5468

**Volume Editors**
Prof. Dr. Niels Pinkwart
Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin, Deutschland
pinkwart (at) informatik.hu-berlin.de
Prof. Dr.-Ing. Johannes Konert
Beuth Hochschule für Technik
Luxemburger Straße 10, 13353 Berlin
jkonert (at) beuth-hochschule.de

**Series Editorial Board**
Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria
(Chairman, mayr@ifit.uni-klu.ac.at)
Torsten Brinda, Universität Duisburg-Essen, Germany
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Infineon, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestädt, Universität Bielefeld, Germany
Wolfgang Karl, KIT Karlsruhe, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Thomas Roth-Berghofer, University of West London, Great Britain
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Andreas Thor, HFT Leipzig, 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 2019
**printed by** Köllen Druck+Verlag GmbH, Bonn




*This book is licensed under a Creative Commons BY-SA 4.0 licence.*

# Vorwort

Wenn Bildungsszenarien zunehmend durch Technologie unterstützt werden, ermöglicht das neben neuen didaktischen Möglichkeiten auch den Abbau von Hürden und Grenzen für den Zugang zum und den Austausch im Bildungssystem. Andererseits können ebendiese Technologien aber auch neue Barrieren schaffen, zu deren Überwindung Lösungen gefragt sind. Teilhabe als Anspruch und als Praxis von Bildung und Wissenschaft bekommt damit eine neue, erweiterte Bedeutung.

Die Tagung der Fachgruppe Bildungstechnologien der Gesellschaft für Informatik e.V. steht im Jahr 2019 unter dem Tagungsmotto *Teilhabe an Bildung und Wissenschaft*. Diese umfasst den Abbau von Zugangshürden, die partizipative Entwicklung von Inhalten oder die Förderung aktiver Teilnahme durch adaptive und personalisierbare Medien. Insbesondere für die Tagung soll der Blick gleichzeitig auf Bildung und Wissenschaft gerichtet werden: hier ermöglichen Medien und Technologien neue Formen von Kollaboration, Integration, Wissenschaftskommunikation sowie neue Verbindungen von Forschung, Lehre und Publikation.

Die DELFI ist die 17. Jahrestagung der Fachgruppe Bildungstechnologien (ehemals Fachgruppe eLearning) der Gesellschaft für Informatik e. V. und findet vom 16. bis 19.September 2019 im Erwin Schrödinger Zentrum der Humboldt-Universität zu Berlin am Campus Adlershof statt, gemeinsam mit der Jahrestagung der Gesellschaft für Medien in den Wissenschaften e.V. (GMW). Diese fruchtbare Zusammenarbeit der Tagungen findet nach Rostock (2005), Berlin (2009), Dresden (2011), München (2015) und Chemnitz (2017) schon zum sechsten Mal statt. Weitergehende Informationen zur Tagung finden sich unter https://fg-bildungstechnologien.gi.de/delfi-tagung/

Die eingereichten Kurz- und Langbeiträge wurden im Doppelblind-Verfahren jeweils von mindestens drei Expertinnen und Experten begutachtet und schließlich bei einem Treffen des Programmkomitees endgültig ausgewählt. Die Einreichenden hatten noch einmal vier Wochen Zeit auf der Basis der Gutachten ihre Beiträge zu finalisieren. Unser herzlicher Dank geht an alle Autorinnen und Autoren sowie an alle Gutachterinnen und Gutachter. Dieser Tagungsband enthält die Beiträge zur DELFI 2019: Insgesamt 13 wissenschaftliche Langbeiträge, davon vier als Vorschlag für den Best Paper Award, sieben Kurzbeiträge, fünf Positionspapiere und sieben Praxisbeiträge. Damit wurden 32 von 69 Beiträgen vom Programmkomitee angenommen. Ergänzt wird das Tagungsprogramm durch 18 Poster-Präsentationen und neun Demos, die jeweils mit einer zweiseitigen Kurzversion im Tagungsband vertreten sind.

Am Montag startet die DELFI & GMW 2019 mit einem Workshoptag, für welchen wir folgende sechs Workshops und ein Tutorial gewinnen konnten:

- *E-Assessment ohne Hürden: Individuelle Vorhaben erfolgreich begleiten und den Umgang mit Heterogenität stärken* (GMW)
- *Inverted Classroom inklusiv gestalten - Potentiale und Grenzen der Digitalisierung* (GMW)

- *Kompetenzen digital: Modellierung, Erfassung, Katalogisierung, Verarbeitung und Zertifizierung* (DELFI)
- *Learning Analytics* (DELFI & GMW)
- *Nutzung von 360°-Video im Kontext forschenden Lernens* (GMW)
- *VR/AR Learning 2019* (DELFI)
- *Wissenschaft richtig gemacht: Tools für Open Scientists und Open Educators* (DELFI Tutorial)

Für die Tagung selbst erwarten wir exzellente eingeladene Keynotes:

- Am Dienstag von Prof. Dr. Petra Grell (TU Darmstadt) zum *Thema Partizipation und Ausgrenzung im Kontext der Digitalisierung*;
- Am Mittwoch von Eva Kern und Julian Reimann (Mitglieder der Zukunfts-AG Digitale ChangeMaker des Hochschulforum Digitalisierung) zum Thema *Digitalisierung und Nachhaltigkeit in der Hochschulbildung zusammen denken* sowie im Anschluss von Stefanie Trzecinski (Geschäftsführerin der KOPF, HAND + FUSS gGmbH Berlin) zum Thema *Digitale Inklusion – ist das machbar?*;
- Am Donnerstag von Ass. Prof. Amy Ogan (Carnegie Mellon University, USA) zum Thema *Designing Culturally-relevant Educational Technology at a Global Scale: Tales from the Field*.

Die Organisation und Durchführung der DELFI & GMW 2019 wurde durch viele Helfende unterstützt. Besonderer Dank geht an dieser Stelle an alle Mitglieder des Programmkomitees für die sehr gewissenhaft angefertigten Gutachten zu den Beiträgen. Darüber hinaus ein herzliches Dankeschön an die vielen, teilweise ehrenamtlichen, Helfenden während der Tagung: studentische Hilfskräfte, Mitarbeiterinnen und Mitarbeiter sowie ehrenamtliche Helfende der Humboldt-Universität zu Berlin, der Beuth Hochschule für Technik Berlin, der Universität Potsdam sowie der Fachhochschule Potsdam.

Zu guter Letzt noch ein Dankeschön an das Präsidium der Humboldt-Universität zu Berlin für die Unterstützung und Bereitstellung der Räumlichkeiten im Erwin Schrödinger Zentrum.

Wir wünschen allen Tagungsteilnehmenden informative und unterhaltsame Beiträge und Diskussionen, die das Forschungsfeld Bildungstechnologien voranbringen und eine schöne Zeit in der Bundeshauptstadt Berlin.

Berlin, im September 2019

Niels Pinkwart
(General Chair)
Humboldt-Universität zu Berlin

Johannes Konert
(Program Chair)
Beuth Hochschule für Technik Berlin

## Tagungsleitung

| | |
|---|---|
| Gesamtleitung: | Niels Pinkwart, Humboldt-Universität zu Berlin |
| Leitung des Programmkomitees: | Johannes Konert, Beuth Hochschule für Technik Berlin |
| Workshops: | Sandra Schulz, Humboldt-Universität zu Berlin |
| Poster & Demos: | Nguyen-Thinh Le, Humboldt-Universität zu Berlin |
| DoktorandInnenkolloqium: | Raphael Zender, Universität Potsdam |

## Programmkomitee

| | |
|---|---|
| Daniel Bodemer | Universität Duisburg-Essen |
| Andreas Breiter | Universität Bremen |
| Torsten Brinda | Universität Duisburg-Essen |
| Ilona Buchem | Beuth Hochschule für Technik Berlin |
| Mohamed Amine Chatti | Universität Duisburg-Essen |
| Matthias Dahlmeyer | Hochschule für Technik und Wirtschaft Berlin |
| Jörg Desel | FernUniversität in Hagen |
| Ralf Dörner | Hochschule RheinMain |
| Hendrik Drachsler | DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation |
| Jens Drummer | Sächsisches Bildungsinstitut |
| Martin Ebner | Technische Universität Graz |
| Wolfgang Effelsberg | Universität Mannheim |
| Albrecht Fortenbacher | Hochschule für Technik und Wirtschaft Berlin |
| Paul Grimm | Hochschule Fulda |
| Joerg Haake | FernUniversität in Hagen |
| Andreas Harrer | Fachhochschule Dortmund |
| Peter Henning | Hochschule Karlsruhe Technik und Wirtschaft |
| Ulrich Hoppe | Universität Duisburg-Essen |
| Dirk Ifenthaler | Universität Mannheim |
| Marco Kalz | Pädagogische Hochschule Heidelberg |
| Paul-Thomas Kandzia | Duale Hochschule Baden-Württemberg |
| Reinhard Keil | Universität Paderborn |
| Andrea Kienle | Fachhochschule Dortmund |
| Ralf Klamma | RWTH Aachen University |
| Johannes Konert | Beuth Hochschule für Technik Berlin |
| Bernd Krämer | FernUniversität in Hagen |
| Detlef Krömker | Goethe Universität Frankfurt am Main |
| Roland Küstermann | Duale Hochschule Baden-Württemberg |
| Nguyen-Thinh Le | Humboldt-Universität zu Berlin |
| Thiemo Leonhardt | Technische Universität Dresden |
| Andreas Lingnau | Hochschule Ruhr West |
| Jörn Loviscach | Fachhochschule Bielefeld |
| Ulrike Lucke | Universität Potsdam |
| Johannes Magenheim | Universität Paderborn |
| Alke Martens | Universität Rostock |
| Christoph Meinel | Hasso-Plattner-Institut |

| | |
|---|---|
| Agathe Merceron | Beuth Hochschule für Technik Berlin |
| Wolfgang Nejdl | L3S und Leibniz Universität Hannover |
| Niels Pinkwart | Humboldt-Universität zu Berlin |
| Hans Pongratz | Technische Universität München |
| Christoph Rensing | Technische Universität Darmstadt |
| Holger Rohland | Technische Universität Dresden |
| Guido Rößling | Technische Universität Darmstadt |
| Nikol Rummel | Ruhr-Universität Bochum |
| Maren Scheffel | Open Universiteit Nederland |
| Daniel Schiffner | DIPF Leibniz-Institut für Bildungsforschung und Bildungsinformation |
| Mareike Schoop | Universität Hohenheim |
| Ulrik Schroeder | RWTH Aachen University |
| Sandra Schulz | Humboldt-Universität zu Berlin |
| Andreas Schwill | Universität Potsdam |
| Sven Strickroth | Universität Potsdam |
| Michael Striewe | Universität Duisburg-Essen |
| Alexander Tillmann | Goethe Universität Frankfurt am Main |
| Stephan Trahasch | Hochschule Offenburg |
| Nicole Wöhrle | Albert-Ludwigs-Universität Freiburg |
| Raphael Zender | Universität Potsdam |

# Organisationsteam

| | |
|---|---|
| Katarzyna Biernacka | Humboldt-Universität zu Berlin |
| Julia Bohle | Fachhochschule Potsdam |
| Sebastian Claus | Humboldt-Universität zu Berlin |
| Florian Fischer | Universität Potsdam |
| Richard Glander | Humboldt-Universität zu Berlin |
| Kristina Herter | Beuth Hochschule für Technik Berlin |
| Stefan Jung | Beuth Hochschule für Technik Berlin |
| Frederic Matthe | Universität Potsdam |
| Daria Padbiarezskaya | Beuth Hochschule für Technik Berlin |
| Michael Rücker | Humboldt-Universität zu Berlin |
| Rudi Rudiak | Humboldt-Universität zu Berlin |
| Julia Ruvela | Universität Potsdam |
| Oskar Schachtschneider | Beuth Hochschule für Technik Berlin |
| Johann Sell | Humboldt-Universität zu Berlin |
| Marlen Schumann | Universität Potsdam |
| Malte Smits | Humboldt-Universität zu Berlin |
| Sara Waldmann | Universität Potsdam |
| Leon Wolff | Fachhochschule Potsdam |

…und weitere Helfende, die nach Drucklegung des Tagungsbandes das Team verstärkt haben. Allen gilt ein herzlicher Dank.

# Inhaltsverzeichnis


## Inklusion

**Alexandra Keus, Svenja Noichl, Ulrik Schroeder**
*Vergleich von Technologien zur Entwicklung barrierefreier mobiler Lernapplikationen* ..... 15

**Svenja Noichl, Susanne Korth, Ulrik Schroeder**
*Konzeption mobiler, inklusiver Aufgaben und Hilfetools in der Ausbildung der Bauindustrie* ..... 21

**Linda Rustemeier**
*Was benötigen Hochschulen, um digitale Barrierefreiheit als Teilhabemöglichkeit umzusetzen?* ..... 27

## Textanalyse

**Niels Seidel, Christian Moritz Rieger**
*Semantic Textual Similarity von textuellen Lernmaterialien und ihre Anwendungen* ... 33

**Tim Steuer, Christoph Rensing**
*Themenübergreifende Diskursklassifikation auf Basis von Word Embeddings und Sequenzfeatures* ..... 45

## Akzeptanz von Digitalen Werkzeugen

**Daniel Biedermann, Hendrik Drachsler**
*Stakeholder attitudes towards digitalization in higher education institutions* ..... 57

**Julian Busse, Aline Lange, Matthias Schumann**
*Effects of Digitalization on Vocational Education and Training: First Results of a Qualitative Study* ..... 67

**Lucas Braeschke, Iris Braun, Felix Kapp, Tenshi Hara**
*Weißt Du es oder rätst Du noch? - ARS mit Rückmeldungen zur Antwortzuversicht bei Lernfragen* ..... 73

**Jana Niemeyer, Alexander Tillmann, Michael Eichhorn**
*Digitalisierungsprozesse an Hochschulen – der Blick der Akteur-Netzwerk-Theorie* ... 85

## Automatisierung & Generierung

**Justin Lefarth, Benjamin Otto, Thilo Schramm, Michael Striewe, Philipp Schmiemann, Michael Goedicke**
*Technische Aspekte der automatischen Aufgabengenerierung für Blended Learning Angebote in der Biologie* ..... 91

**Robert Garmann**
*Ein Format für Bewertungsvorschriften in automatisiert bewertbaren Programmieraufgaben* ..... 103

**Michael Striewe**
*Automatische Aufgabengenerierung über Linked Open Data am Beispiel der Archäologie* ..... 115

**Miriam Weigel, Tatyana Podgayetskaya, Reinhold Hübl**
*Digitalisiertes Feedback für verschiedene Frageformate von Online-Übungsaufgaben* ..... 121

**Patrick Palsbröker, Matthias Lutz, Carsten Gips, Matthias König**
*Entwicklung eines modernen Lehr- und Lernkonzepts für Maschinelles Lernen mit Infrastructure-as-a-Service und Cloud Computing* ..... 127

## VR/AR/3D

**Axel Wiepke, Eric Richter, Raphael Zender, Dirk Richter**
*Einsatz von Virtual Reality zum Aufbau von Klassenmanagement-Kompetenzen im Lehramtsstudium* ..... 133

**Anna Klingauf, Johannes Funk, Angela Lüüs, Ludger Schmidt**
*Wirkung von interaktiven 3D-360°-Lernvideos in der praktischen Ausbildung von Handwerkern* ..... 145

**Lucas Küntzer, Michael Hoffmann, Georg Rock**
*"Virtual Reality as an E-Learning tool in the student-centered vehicle development project "proTRon"* ..... 157

## Adaptation

**Mathias Bauer, Jacqueline Schuldt, Heidi Krömker, Beatrix Bau, Marie Jeanne Webers**
*Förderung der Lernmotivation durch adaptives E-Learning: Komparative Evaluation von Techniken zur adaptiven Nutzerführung* ..... 163

**Sebastian Wollny, Jan Schneider, Marc Rittberger, Hendrik Drachsler**
*"What can I help you with today?" - Exploring Opportunities of Learner Modeling for Online Educational Portals* ..... 175

**Hendrik Geßner, Alexander Kiy**
*A mobile campus application as a sensor node for Personal Learning Environments* 187

**Albrecht Fortenbacher, Manuel Ninaus, Haeseon Yun, René Helbig, Korbinian Moeller**
*Sensor Based Adaptive Learning - Lessons Learned* ..... 193

**Haeseon Yun, Shirin Riazy, Albrecht Fortenbacher, Katharina Simbeck**
*Code of Practice for Sensor-Based Learning* ..... 199

## Recht & Ethik

**Jens Lechtenbörger**
*Simplifying license attribution for OER with emacs-reveal* ..... 205

**Marc Sauer, Manuel Froitzheim, Andreas Hoffmann**
*Praktische Erfahrungen mit einem elektronischen Prüfungsprozess-Management für eAssessments* ..... 217

**Shirin Riazy, Katharina Simbeck**
*Predictive Algorithms in Learning Analytics and their Fairness* ..... 223

**Christoph Richter, Heidrun Allert**
*Towards a Critical Design Agenda in Support of Collective Learning Ecologies* ..... 229

## Games / Gamification

**Elias John**
*PLAY & CODE – Eine Serious Game basierte Lernplattform* ..... 235

**Heinrich Söbke**
*Die Self-Determination Theory als Rahmen für Deep Gamification in didaktischen Kontexten* ..... 247

**Sven-Bjarne Seiffert, Andreas Lingnau**
*Interaktive Unterstützung bei der Begehung eines Kunstlehrpfads: Die WaldSkulpturenWeg-App* ..... 253

## Interaktives Lernen

**Sebastian Hobert**
*How Are You, Chatbot? Evaluating Chatbots in Educational Settings – Results of a Literature Review* ..... 259

**Sara Durski, Marguerite Schwarzbach, Lisa Frey, Yasmin Ashimi, Ute Massler, Wolfgang Müller**
*Leseflüssigkeitstraining mit Amazon Alexa* ..... 271

**Anja Pfennig, Jörg Maier-Rothe**
*Lehrfilme einfach einfach machen – erfolgreiche Konzeptionierung von Peer-to-Peer Lehrfilmen* ..... 277

## Poster & Demos

**Anika Westburg**
*An E-Learning Tool for Parsing-as-Deduction* ..... 283

**Tobias Scheidat, Sandra Scholz, Mailyn Engel, Stephanie Böhnke, Heike Mrech, Korinna Bade**
*Die intoMINT-App: Spielerisch für MINT begeistern und Berufsbilder vermitteln* ..... 285

**René Röpke, Matthias Ehlenz, Thiemo Leonhardt, Ulrik Schroeder**
*Dungeons & DFAs* ..... 287

**Tobias Moebert, Ulrike Lucke**
*E.V.A. – Emotionen Verstehen und Ausdrücken* ..... 289

**Nadine Schröder**
*EduArc – eine verteilte Infrastruktur für OER* ..... 291

**Ala Naoufal, Johannes Konert, Hicham Naoufal**
*EDUSPACE – Social Media Lernplattform mit Schwerpunkt Inklusion* ..... 293

**Mathias Magdowski**
*Eine „umgedrehte" Video-Nachbesprechung einer Leistungskontrolle im E-Technik-Grundstudium* ..... 295

**Sebastian Hobert, Florian Berens**
*Einsatz von Chatbot-basierten Lernsystemen in der Hochschullehre - Einblicke in die Implementierung zweier Pedagogical Conversational Agents* ..... 297

**Sima Darabi**
*Entwicklung eines E-Assessments basierend auf sequentiellem Testen für die sechsstufige Zensierung* ..... 299

**Eric Müller, Franz Schwarzer, Karsten Weicker**
*Geocatch – Upsizing von geometrischen Konstruktionsaufgaben mithilfe einer Android-App* ..... 301

**Thiemo Leonhardt, Matthias Ehlenz**
*Greifbares Lernen von Informatikinhalten - Aktive Tangibles auf kapazitiven Multitouch-Tischen* ..... 303

**Raphael Zender, Matthias Weise, Ulrike Lucke**
*HandLeVR - Handlungsorientiertes Lernen in der VR-Lackierwerkstatt* ..... 305

**Sven Judel, René Röpke, Nadja Zaric, Ulrik Schroeder**
*ILSA – ein integriertes Lernstil-Analysesystem für Moodle* ..... 307

**Svenja Noichl, Ulrik Schroeder**
*Informatik Grundbildung im Alter – Ein eLearning Konzept* ..... 309

**Yasmin Patzer, Niels Pinkwart**
*Inklusives Lernen mit barrierfreien Videos - Ein Fallbeispiel* ..... 311

**Manuel Froitzheim, Michael Schuhen**
*Interaktive kollaborative Elemente aus einer domänenspezifischen Perspektive des Wirtschaftsunterrichts* ..... 313

**René Röpke, Kathrin Larisch, Sven Schöbel, Ulrik Schroeder**
*Mit der Lupe unterwegs – eine spiel-basierte Lernanwendung zu Sicherheit im Internet* ..... 315

**René Röpke, Ulrik Schroeder**
*Mit G/P/S durch die Welt der spielbasierten Lernanwendungen und Serious Games für IT-Sicherheit* ..... 317

**Thiemo Leonhardt, Matthias Ehlenz, René Röpke, Nadine Bergner, Ulrik Schroeder**
*MTLG – Helping students and teachers creating and analysing simulations and games from a didactical and technical perspective* ..... 319

**Uwe Lorenz, Tatjana Reinstorf**
*Nachhaltiges und Aktivierendes Lernen durch eigene Erklärungen (NALE) mit geteilten, digitalen Portfolios* ..... 321

**Anke Pfeiffer, Dieter Uckelmann**
*Open Digital Lab for You (DigiLab4U) – Entwicklung und Pilotierung eines mediendidaktischen Designs für laborbasiertes Lehren und Lernen* ..... 323

**Mareike Bartels, Eileen Lübcke, Gabi Reimann, Finja Huckfeldt, Ivo van den Berk,**
*Verbesserung der Lehrpraxis durch die Systematisierung erprobter Lösungen: Die Plattform Patternpool* ..... 325

**Till Massing, Natalie Reckmann, Benjamin Otto, Alexander Blasberg, Christoph Hanck, Michael Goedicke**
*Quantitative Methodenkompetenzen im Praxissemester in Lehramtsstudiengängen*... 327

**Bonny Brandenburger**
*Promoting Competencies of the Future – The Role of University-integrated Makerspaces* ..... 329

**Robert Meyer, Juliane Aulinger**
*UnterrichtOnline.org - ein Videografieportal für alle Phasen der LehrerInnenbildung – Interaktives und kollaboratives webbasiertes Lehren und Lernen mit Unterrichtsvideos* ..... 331

**Jan Bernoth, Alexander Kiy, Ulrike Lucke**
*Analytics-Dashboard für Lehre, Forschung und Transfer*..... 333

**Anja Hawlitschek, André Dietrich, Sebastian Zug**
*Welche inhaltliche Unterstützung ist für Studierende beim Programmieren in einem Remote-Labor hilfreich?* ..... 335

Verzeichnis der Autorinnen und Autoren ..... 337

# Vergleich von Technologien zur Entwicklung barrierefreier mobiler Lernapplikationen

**Web-Apps, Cross-Plattform-Apps und Native Android Apps im Vergleich**

Alexandra Keus[1], Svenja Noichl[2] und Ulrik Schroeder[2]

**Abstract:** Während durch die Digitalisierung die Zahl der Smartphone- und Tablet-Nutzerinnen und Nutzer stetig steigt, leben gleichzeitig zahlreiche Menschen mit Behinderungen in Deutschland. In Bezug auf mobile Lehr- und Lernangebote wird das Thema Barrierefreiheit immer bedeutender. Mit Hilfe von manuellen und toolgestützten Tests wird eine beispielhafte Lernapplikation, zur Anwendung erlernten Interaktionsmöglichkeiten mit mobilen Geräten, auf Barrierefreiheit überprüft. Dabei wird die App in drei Varianten, einer nativen Android App, einer Cross-Plattform-App sowie einer Web-App betrachtet. Es zeigt sich, dass barrierefreie Entwicklung gemäß existierender Richtlinien grundsätzlich möglich ist, jedoch alle App-Varianten gewisse Vor- und Nachteile bieten.

**Keywords:** Barrierefreiheit, Cross-Plattform-App, Native Android App, Web-App, Lernapplikation

## 1 Einleitung und Motivation

Die Anzahl der Smartphone-Nutzerinnen und Nutzer ist in allen Altersstufen in den vergangenen Jahren immer weiter gestiegen. Zusätzlich spielen mobile Applikationen durch die zunehmende Digitalisierung in allen Lebensbereichen eine immer größere Rolle. Bei der Entwicklung sind die unterschiedlichen Betriebssysteme mobiler Geräte zu berücksichtigen. Um möglichst alle potentiellen Nutzerinnen und Nutzer erreichen zu können, ist zumindest die Entwicklung einer Android-App und einer iOS-App notwendig. Alternativ kann eine Cross-Plattform-App entwickelt werden. Das Ziel hierbei ist es eine App zu entwickeln, welche auf unterschiedlichen Plattformen nutzbar ist. Gleichzeitig gibt es laut Angaben des Statistischen Bundesamtes Ende 2017 ca. „7,8 Millionen schwerbehinderte Menschen in Deutschland“ [St19]. Eingeschlossen werden hier nur Personen mit einem Behinderungsgrad von mindestens 50% [St19]. Nicht nur für diese Nutzerinnen und Nutzer, sondern auch für Menschen mit geringeren Einschränkungen, in beispielsweise der Sehkraft, können Probleme in der Nutzung von mobilen Applikationen auftreten.

Detailliertere Darstellungen des Konzepts der für die Tests genutzten Beispiel-Applikation sowie der Testergebnisse sind in [Ke19] zu finden.

[1] RWTH Aachen, Ahornstraße 55, 52074 Aachen, alexandra.keus@rwth-aachen.de
[2] RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, {noichl, schroeder}@informatik.rwth-aachen.de

doi:10.18420/delfi2019_247

## 2 Zugrundeliegende Technologien und Richtlinien

Neben der nativen Android-App wird ein Cross-Plattform-Ansatz betrachtet. Nach einer Statistik von appfigures insights war Cordova Ende 2017 das am häufigsten verwendete Framework zur Entwicklung von nicht nativen Apps [Ap18]. Die Applikationen werden mithilfe von Webtechnologien, wie HTML, CSS und JavaScript, entwickelt. Hier wurde zusätzlich Ionic verwendet, um ein möglichst natives Design zu erreichen. Durch die Verwendung eines auf Webtechnologien basierenden Frameworks, können auch die Richtlinien für barrierefreie Webentwicklung gut berücksichtigt werden, da aus technischer Sicht zunächst eine Webseite entwickelt wird. Neben der Nutzung als Webseite im Browser, kann der Code mit Hilfe von Cordova auch als native App für Android, iOS und Windows zur Verfügung gestellt werden. Als Grundlage für Richtlinien für barrierefreie Webentwicklung wurden die Web Content Accessibility Guidelines (WCAG 2.0) [Ca08], die Barrierefreie-Informationstechnik Verordnung (BITV 2.0) [Bu11], die WCAG 2.1 [Wc18], das Universal Design [Di19] sowie die Do's and Dont's der britischen Regierung [Go17] berücksichtigt.

## 3 Vorgehen und Fragestellungen der durchgeführten Tests

Mithilfe von manuellen und toolgestützten Tests soll die Frage beantwortet werden, inwieweit mobile Applikationen, die mit aktuell gängigen Technologien, basierend auf Richtlinien zur barrierefreien Webentwicklung, entwickelt wurden, in der Praxis barrierefrei sind. Die Nutzung externer Komponenten (z. B. Braillezeile, Schalterzugriff, Tastatur) wurde nicht betrachtet. Gemäß des Leitfadens für Android Entwickler, wurden die Apps im Rahmen des manuellen Testens auf ihre Nutzbarkeit mittels eines Screenreaders, den Einfluss einer Anpassung von Anzeige- und Schriftgröße sowie die Anpassbarkeit von Kontrast- und Farboptionen ohne funktionale Einschränkungen hin überprüft [An19]. Ziel ist die Evaluation des Verhaltens der Apps bei deren Nutzung. Zur Ermittlung der Nutzbarkeit der Apps mittels eines Screenreaders wurde der Screenreader-Dienst *TalkBack* verwendet. *TalkBack* wird von Android angeboten und kann über die Android Accessibility Suite[3] installiert werden. Für die toolgestützen Tests wurden die drei Apps einem Test mit dem Android Accessibility Scanner[4] unterzogen. Für Android Geräte mit einer Android-Version von mindestens 6.0 kann dieses Tool verwendet werden um Apps auf dem Gerät auf ihre Barrierefreiheit hin zu untersuchen. Alle Tests wurden auf einem Samsung Galaxy Tab S3 mit Android-Version 8.0, API 26 und einer Displayauflösung von 2048x1536 Pixel durchgeführt. Um die Web-App auf demselben Gerät testen zu können, wurde diese über einen Raspberry Pi im lokalen Netz gehostet, der Zugriff darauf erfolgt über den Browser Chrome. Abb. 1 zeigt Screenshots der App mit ihren unterschiedlichen Interaktionsmöglichkeiten.

[3] https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback&hl=de
[4] https://support.google.com/accessibility/android/faq/6376582

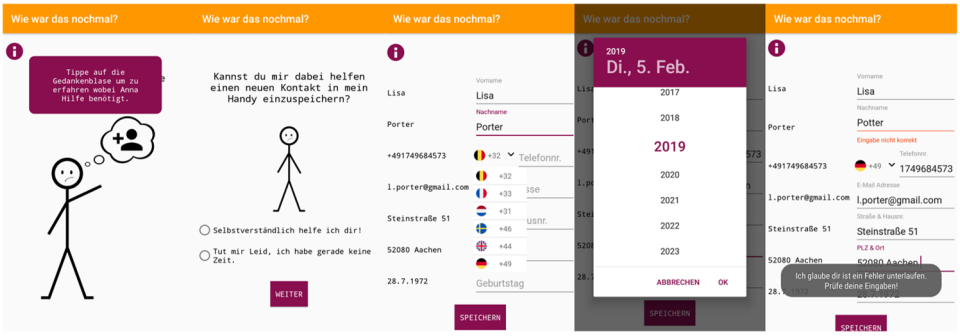



Abb. 1: Screenshots der App mit unterschiedlichen Interaktionsmöglichkeiten (vgl. Ke19)

# 4 Ergebnisse der Manuellen und Toolgestützten Tests

Bei den folgenden Ergebnissen der Tests ist zu beachten, dass die Ergebnisse der Web-App sowie die der Cross-Plattform-App übereinstimmen, sofern keine expliziten Unterschiede genannt werden.

## 4.1 Ergebnisse des Manuellen Tests

Bei den manuellen Tests unter Verwendung des Screenreaders TalkBack wurden sieben Kategorien betrachtet (siehe Tabelle 1). Die native Android App schneidet bei diesem Test weitestgehend gut ab. Das größte Problem trat hier bei fehlerhaften Eingaben in Eingabefeldern auf. Die daraus resultierenden Fehlermeldungen werden nur einmalig vorgelesen und bei erneutem Vorlesen des betroffenen Feldes nicht wieder berücksichtigt. Dies widerspricht dem Gedanken, dass Fehlermeldungen unterhalb von Eingabefeldern zum leichteren Erkennen und Beheben von Fehlern durch die Nutzerinnen und Nutzer beitragen sollen. Fehlermeldungen sind somit nur für Personen hilfreich, die nicht auf die Verwendung eines Screenreaders angewiesen sind. Beim Date-Time-Picker wird bei Auswahl von Monat und Tag erst beim Antippen des darunterliegenden Kalenderelements die entsprechende Auswahl vorgelesen und nicht bereits beim Blättern durch z. B. die Monate. Kleine Einschränkungen konnten bei Info-Popups und Dialogen festgestellt werden. Die enthaltenen Texte wurden hier teilweise zweimal hintereinander vorgelesen. Bei Web-App und Cross-Plattform-App werden Fehlermeldungen nicht automatisch vorgelesen, dafür kann diese jedoch angewählt werden, indem über die Nachricht unterhalb des betroffenen Feldes gestrichen wird. Ein wiederholtes Vorlesen ist somit möglich. Optimal wäre hier eine Kombination der Funktionalitäten aus initialem Vorlesen der Fehlermeldung, wie in der nativen App und wiederholbarem Vorlesen. Der Date-Time-Picker wird in diesen App Versionen nicht als Kalenderansicht, sondern als scrollbare Spalten dargestellt. Während das scrollen die Nutzung im Screenreader-Modus erschwert, da ein leichtes „verblättern“ möglich ist, bietet dieses Variante den Vorteil, dass das jeweils gewählte

Datum stets vorgelesen wird. Die Inhalte von Info-Popups und Dialogen werden nicht automatisch vorgelesen. Somit ist für Nutzerinnen und Nutzer mit Sehbeeinträchtigungen unklar, ob durch die Interaktion mit einer Schaltfläche eine Aktion ausgeführt wurde. Um dies zu verifizieren, müssen die Nutzerinnen und Nutzer schlimmstenfalls einmal über den gesamten Bildschirm fahren. Auch Toasts (kurze Mitteilungen in der App) werden nicht automatisch vorgelesen, wodurch für Nutzerinnen und Nutzer mit Sehbeeinträchtigung eine typische App-Funktionalität verloren geht. Bei Radiobuttons wird bei einem Wechsel der Auswahl „nicht aktiviert“ gefolgt von „aktiviert“ vorgelesen. Dieses Verhalten des Screenreaders macht die Bedienung schwer nachvollziehbar. Weiterhin wird das Dropdown-Menü nicht als ein Element, sondern als drei einzelne Elemente behandelt, was die Interaktion mit diesem Auswahlbereich teilweise erschwert. Beim Testen der Web-App ist zusätzlich zu beachten, dass sichergestellt werden muss, dass die Spracheinstellung der App mit der von *TalkBack* verwendeten Sprache übereinstimmt. Andernfalls könnte dies beispielsweise zu einer englischen Aussprache von deutschen Texten führen. Im Vergleich zur nativen App fällt außerdem auf, dass einige der Textbereiche kleiner, also schwerer anzuwählen sind als in der nativen App. Tabelle 1 zeigt wie gut die drei App-Varianten bei der Analyse mittels des Screenreaders *TalkBack* abgeschnitten haben.

| | Android App | Cross-Plattform-App + Web-App |
|---|---|---|
| Texte, Beschriftungen | ✓ | ✓ |
| Radiobuttons, Checkboxen | ✓ | (✓) |
| Eingabefelder | ✓ | ✓ |
| Fehlermeldungen | (✓) | (✓) |
| Toasts | ✓ | (✓) |
| Date-Time-Picker | (✓) | (✓) |
| Info-Popups, Dialoge | (✓) | ✖ |

Tabelle 1: Ergebnisse des Tests mit TalkBack

Bei Anpassungen der Schriftgröße in den Geräteeinstellungen werden die Schriftgrößen bei der nativen Android App und Cross-Plattform-App entsprechend angepasst. Bei der Web-App führt diese Anpassung jedoch nicht zu einer Änderung der Schriftgröße. Die Android App konnte ohne Einschränkung mit Vergrößerungen bis zur höchsten von sieben Stufen umgehen, bei der Cross-Plattform-App traten ab Stufe fünf Probleme beim Date-Time-Picker auf. Eine Anpassung der Kontrastverhältnisse über die Geräte Einstellungen führte in keiner Version zu Problemen.

### 4.2 Ergebnisse des Toolgestützten Tests

Beim toolgestützten Test der nativen App mit dem Accessibility Scanner spiegelten sich einige Aspekte des manuellen Tests wieder. Der Elementtyp TextInputLayout, jene Eingabefelder, bei denen Fehlermeldungen bei nicht korrekt ausgefüllten Eingaben unter diesen angezeigt werden, wurde als möglicherweise problematisch eingestuft. Es wird angegeben, dass Bedienungshilfen diesen Elementtyp möglicherweise nicht auflösen können. Als weiteres kritisches Element wird der Date-Time-Picker angemerkt. Demnach ist der

Berührungszielbereich für die Auswahl des Jahres und des Datums zu klein. Weiterhin wird angemerkt, dass es sich bei der Kalenderansicht um einen nicht unterstützten Elementtyp handelt. Einer der auffälligsten besteht im Kontrastverhältnis des App-Titels im Header. Bei den hier verwendeten Farben liegt das Kontrastverhältnis bei 2,2:1. Ein Kontrastverhältnis von mindestens 3:1 wird vorgesehen. Da dieser Aspekt jedoch hauptsächlich von dem gewählten Farbschema abhängt, wird dies hier im Folgenden nicht weiter berücksichtigt, wenngleich dieser Aspekt natürlich bei der Entwicklung beachtet werden sollte. Ein weiterer Kritikpunkt ist an dieser Stelle außerdem, dass das Kontrastverhältnis bei der Fehlermeldung in Bezug auf den Hintergrund mit 3:1 zu gering gewählt sind. Aufgrund der geringen Textgröße wird hier ein Kontrastverhältnis von mindestens 4,5:1 empfohlen. Tests mit anderen Signalfarben an dieser Stelle führten jedoch zu ähnlichen Fehlern. Hier bleibt abzuwägen, ob das Kontrastverhältnis eingehalten werden sollte, oder mittels einer Signalfarbe deutlich auf einen Fehler hingewiesen werden sollte. Bei den Tests der Web-App und der Cross-Plattform-App wurden vom Accessibility Scanner mehr Kritikpunkte angeführt. Für die Umsetzung der Web-App und der Cross-Plattform-App wurde die in den Vorgaben für barrierefreie Webseiten angegebene Mindestgröße von 44x44px für anklickbare Flächen wie Buttons und Eingabefelder verwendet. Dies resultiert bezogen auf das genutzte Testgerät, laut dem Accessibility Scanner, bei der Web-App in generelle zu kleine Berührungszielbereiche. Widersprüchlich sind hier die Auswertungen der Kontrastverhältnisse. Während das Kontrastverhältnis bei einem Button als zu gering angemerkt wird, führt dieselbe Farbwahl bei Titel und Header der Seite nicht zu einem Warnhinweis. Auch die Größe von Radiobuttons wird als zu klein angemerkt. Allerdings ist hier zu beachten, dass die Kritik durch den Scanner hier lediglich auf der Größe der runden Markierungsfelder basiert. Der Radiobutton lässt sich jedoch zusätzlich durch antippen des dazugehörigen Auswahltextes aktivieren. Somit ist der tatsächliche Berührungszielbereich des Radiobuttons größer als vom Accessibility Scanner erkannt. Genauso wie bei der nativen App führt auch bei der Web-App der Date-Time-Picker zu großen Problemen. Neben zu kleinen Zielbereichen konnten zahlreiche weitere Verbesserungsvorschläge entdeckt werden. Insgesamt wurden 302 Vorschläge in Bezug auf den Date-Time-Picker angegeben. Nach genauerer Betrachtung scheint dies daran zu liegen, dass der Scanner an dieser Stelle nicht in der Lage ist das Element richtig auszuwerten. Interessant ist, dass es, anders als bei der nativen Android App, keine Kritik wegen eines zu geringen Kontrasts in Bezug auf die Textfarbe der dargestellten Fehlermeldung bei fehlerhaften Eingaben gibt. Der Kontrast liegt auch hier unter 4,5:1. Theoretisch müsste der Scanner hier also, wie auch bei der nativen App, einen Verbesserungsvorschlag anbieten.

# 5 Ausblick

Der Fokus lag hier insbesondere auf Aspekten der Barrierefreiheit, welche für den Einsatz in mobilen Lernapplikationen relevant sind. An der hier klein gehaltenen Beispiel-Implementierung einer App zur Anwendung von erlerntem Wissen über die Interaktion mit mobilen Geräten, konnte gezeigt werden, dass die Entwicklung barrierefreier mobiler Lernapplikationen grundsätzlich möglich ist. Es stellte sich jedoch auch heraus, dass es

nicht den „perfekten“ Ansatz für die Umsetzung solcher Projekte gibt. Aufgrund der hier dargestellten Ergebnisse wären neben ähnlichen Untersuchungen in größeren Projekten auch weitere Tests mit anderen Cross-Plattform-Frameworks denkbar. Da alle hier beschriebenen Varianten über Vor- und Nachteile verfügen wäre auch eine weitere Untersuchung dahingehend interessant, ob die jeweiligen Komponenten aus den unterschiedlichen App-Varianten in Kombination verwendet werden könnten. Als abschließendes Fazit lässt sich sagen, dass barrierefreie, mobile Lernapplikationen für viele Lernenden mit diversen Beeinträchtigungen, gerade in der heute immer mehr digitalisierten Welt von besonderer Relevanz sind. An einheitlichen Richtlinien für unterschiedliche Umsetzungsvarianten mobiler Applikationen mangelt es jedoch noch, wodurch ein optimales Lernerlebnis für Lernende mit Beeinträchtigungen nicht unbedingt gegeben ist.

Literaturverzeichnis

[An19] Android Developers, Test your app’s accessibility https://developer.android.com/training/accessibility/testing, Stand: 14.03.2019.

[Ap18] appfigures, iOS Developers Ship 29% Fewer Apps In 2017, The First Ever Decline – And More Trends To Watch, https://blog.appfigures.com/ios-developers-ship-less-apps-for-first-time/, Stand: 14.03.2019.

[Bu11] Bundesministerium der Justiz und für Verbraucherschutz, Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung – BITV 2.0), https://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html, Stand: 14.03.2019.

[Ca08] Caldwell, B.; Cooper, M.; Reid, L. G.; Vanderheiden, G.: Web content accessibility guidelines (WCAG) 2.0. WWW Consortium (W3C). 2008.

[Di19] Digital informiert im Job integriert, Universelles Design und Design für Alle, http://www.di-ji.de/index.php?option=com_content&view=article&id=45, Stand: 14.03.2019.

[Go17] GOV.UK, Accessibility in government, https://accessibility.blog.gov.uk/2017/12/18/what-working-on-gov-uk-navigation-taught-us-about-accessibility/, Stand: 14.03.2019.

[Ke19] Keus, A.: Barrierefreie App Entwicklung – Native, Plattformübergreifende & Web-Apps im Vergleich. Bachelorarbeit, RWTH Aachen. 2019.

[St19] Statistisches Bundesamt, https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Gesundheit/Behinderte/BehinderteMenschen.html, Stand: 14.03.2019.

[Wc18] W3C, Web Content Accessibility Guidelines (WCAG) 2.1, https://www.w3.org/TR/WCAG21/, Stand: 14.03.2019.

# Konzeption mobiler, inklusiver Aufgaben und Hilfetools in der Ausbildung der Bauindustrie

Svenja Noichl[1], Susanne Korth[2] und Ulrik Schroeder[1]

**Abstract:** Lernende mit sonderpädagogischem Förderbedarf oder persistierenden Lernstörungen können in Schule und Ausbildung vor besonderen Herausforderungen stehen. Am Beispiel der Ausbildung in der Bauindustrie wird ein Konzept zur Unterstützung von Auszubildenden entwickelt. Am konkreten Fallbeispiel des Mauerwerksbaus sollen mobile Lernapplikationen mithilfe eines Cross-Plattform-Ansatzes entwickelt werden. Innerhalb der Aufgaben, die in der Lernapplikation bearbeitet werden können, werden Hilfetools ansetzen, um die Lernenden bei der Lösung der Aufgaben zu unterstützen. Im Sinne der Inklusion sollen individuelle Hilfestellungen möglich sein und Richtlinien zur barrierefreien Entwicklung berücksichtigt werden.

**Keywords:** berufliche Ausbildung, Inklusion, Lernbarrieren, Cross-Plattform-Apps

## 1 Einleitung

Im Juni 2018 veröffentlichte die Kultusministerkonferenz die Dokumentation „Sonderpädagogische Förderung in Schulen 2007 bis 2016“ [KMK18]. Im Zwei-Jahres-Rhythmus werden an dieser Stelle die Zahlen zu Schülerinnen und Schülern, welche eine sonderpädagogische Förderung benötigen, gesammelt. Von 523.000 Schülerinnen und Schülern, die 2016 mit sonderpädagogischer Förderung unterrichtet wurden, können 36,5% dem Förderschwerpunkt Lernen zugeordnet werden. [KMK18]

Die Schwierigkeiten betroffener Schülerinnen und Schüler bestehen auch in ihrem weiteren Werdegang fort. So können auch in der Ausbildung der Bauindustrie ausgeprägte Lernbeeinträchtigungen bei Auszubildenden beobachtet werden. Dies betrifft sowohl Auszubildende mit zuvor festgestelltem sonderpädagogischen Förderbedarf im Förderschwerpunkt Lernen als auch solche, bei denen in ihrer Schullaufbahn ein derartiger Förderbedarf zwar nicht diagnostiziert wurde, die aber dennoch persistierende Lernstörungen z. B. im Lesen, Schreiben und/oder Rechnen aufweisen. Hier können beispielsweise Schwierigkeiten beim Verstehen von Aufgabenstellungen durch unbekannte Fachtermini oder beim Transfer von Rechenregeln auf neue Aufgabenstellungen auftreten. An dieser Stelle wollen wir in dem Projekt MeinBerufBau ansetzen. Das Verbundprojekt wird durch das BMBF aus den Mitteln des ESF Deutschland im Rahmen des Programms „Inklusion durch digitale Medien in der beruflichen Bildung“ gefördert. Ziel des Vorhabens ist es, mithilfe inklusiver Aufgaben

[1] RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, {noichl, schroeder}@informatik.rwth-aachen.de
[2] RWTH Aachen, Lehr- und Forschungsgebiet Fachdidaktik Bautechnik, Mies-van-der-Rohe Straße 1, 52074 Aachen, korth@fdb.rwth-aachen.de

doi:10.18420/delfi2019_243

und Hilfetools, welche in barrierefreien, mobilen Lernapplikationen eingebunden werden, mögliche Barrieren abzubauen. Durch die Hilfetools soll das Lösen der Aufgaben innerhalb der mobilen Lernapplikationen unterstützt werden. Gleichzeitig werden bei der Entwicklung Richtlinien zur barrierefreien Entwicklung berücksichtigt. Auf diese Weise können die Lehr- und Lernprozesse in der Ausbildung der Bauindustrie begünstigt werden.

Dieses Paper zeigt in Kapitel zwei, auf welche Art und Weise die Aufgaben aktuell, ohne technische Unterstützung, vermittelt werden. Folgend werden exemplarisch Aufgabenbeispiele des Bereichs *Mauerwerksbau* und die bei der Bearbeitung dieser Aufgaben identifizierten Barrieren dargestellt. In Kapitel drei wird anschließend ein konzeptioneller Ansatz zum Abbau dieser Barrieren mithilfe einer mobilen Anwendung vorgestellt. Letztlich werden in Kapitel vier die vorgesehenen technischen Aspekte sowie die geplante Einbettung der Anwendung in den Ausbildungskontext diskutiert.

## 2 State of the Art

Im Rahmen der beruflichen Ausbildung in der Bauindustrie besuchen die Auszubildenden des 1. Ausbildungslehrjahres in der beruflichen Grundbildung Lehrgänge verschiedener berufsspezifischer Fachbereiche. Eine Befragung von 12 Ausbildern der unterschiedlichen Bereiche der Bauindustrie am Ausbildungszentrum Kerpen[3] (kurz: ABZ Kerpen) zeigte, dass wie erwartet ein hoher Praxisanteil innerhalb dieser Lehrgänge gegeben ist. So ist beispielsweise im Bereich Mauerwerksbau eine erste Aufgabenstellung, eine gerade Wandscheibe entsprechend der gegebenen Rahmenbedingungen zu mauern. Neben dem fachgerechten Umgang mit Messwerkzeugen ist dabei auch die rechnerische Kontrolle der Mauermaße ein Handlungsschritt, der sorgfältig und fehlerfrei ausgeführt werden muss, um dem Ziel des Qualitätsanspruchs bei der Produktherstellung gerecht zu werden. Die Aufgabenstellung liegt den Auszubildenden meist in Papierform oder als Tafelanschrieb vor und bietet somit keine weiteren Hilfestellungen bei besonderen Einschränkungen oder individuellen Problemen.

Im Mauerwerksbau ist, gemäß dem Lehrbuch *Handelnd Lernen in der Bauwirtschaft* des Bundesinstituts für Berufsbildung (kurz: BiBB) [Fi06], eine erste mögliche Projektaufgabe das Mauern der Wände eines Ferienhauses. Im Rahmen dieser Projektaufgabe müssen dann in Teilaufgaben die einzelnen Teile des Mauerwerks errichtet werden. Gegeben sind bei einer solchen Aufgabenstellung, neben einer kurzen Aufgabenbeschreibung, unterschiedliche Ansichten (Vor-, Drauf-, Seitenansicht) der Mauer, aus denen die benötigten Steine und Fugen abgelesen werden können. Nach einer Informationsphase, bei der die Auszubildenden die Informationen zur Teilaufgabe in Form von Aufgabenblättern oder Tafelanschrieben erhalten, folgt eine Planungsphase. In dieser sollen die Auszubildenden selbstständig unter anderem das benötigte Material für den Bau der Mauer planen. Für diesen Schritt ist es wichtig, dass die Auszubildenden anhand der

[3] https://www.abz-kerpen.de/

ihnen zur Verfügung stehenden Aufgabeninformationen, die richtigen Materialien wählen und die gegebenen Maße abgleichen können. Beispielsweise wäre es hier sinnvoll, zu berechnen, welche Maße eine Mauer mit den selbst ermittelten Materialien haben wird, und ob dieses Maß mit dem der Aufgabenstellung übereinstimmt, um so die eigenen Ergebnisse bei dieser Teilaufgabe überprüfen zu können. [Fi06]

Es gibt unterschiedliche Lernbarrieren, die bei der Bearbeitung solcher Aufgaben hinderlich sein können. Bereits die Aufgabenstellung erweist sich als ein Schwerpunkt. Das fehlende Verständnis oder Verwechselungen von verwendeten Fachbegriffen können hier Anknüpfungspunkte für Hilfetools (siehe Kapitel 3) bieten. Auch Schwierigkeiten bei der späteren Aufgabenbearbeitung aufgrund nicht vollständig gelesener Informationsanlagen (Tabellen, technische Zeichnungen) sowie eventuelle Leseschwächen können dazu führen, dass die Aufgabenstellung nicht richtig verstanden wird. Diese kognitiven Lernbarrieren können im weiteren Arbeitsprozess dazu führen, dass die Auszubildenden nicht die geforderten Werte, wie z. B. das Außenmaß[4] der Mauer, berechnen bzw. berechnen können. Die fachgerechte Kontrolle der Mauermaße ist damit nicht durchführbar. Selbst die Anzahl von Steinen und Fugen sind aufgrund von den genannten Lernbarrieren möglicherweise nicht ermittelbar.

## 3 Konzeption der Aufgaben und Hilfetools

Das im Folgenden vorgestellte Konzept für Aufgaben und Hilfetools sollen an den in Kapitel 2 aufgeführten Lernbarrieren ansetzen und eine technische Unterstützung bei der Bearbeitung bieten. Gleichzeitig sollen die Aufgaben und Hilfetools in einem Gesamtkonzept in die unterschiedlichen Bereiche der Bauindustrie eingegliedert werden. Das bedeutet, das Gesamtkonzept der App beinhaltet mehrere Bereiche der Bauindustrie. In einem dieser Bereiche gibt es eine Anzahl verschiedener Aufgabentypen. Für einen Aufgabentyp wiederum sollen konkrete Aufgaben teilweise automatisch generiert werden können. Innerhalb einer konkreten Aufgabe bieten dann die Hilfetools Unterstützung bei der Bearbeitung der Aufgabe. Abbildung 1 dient zur Verdeutlichung der Zusammenhänge der zuvor beschriebenen Bestandteile.

[4] Außenmaß oder auch Pfeilermaß bezeichnet im Mauerwerksbau die Länge einer Mauer (vgl. [Ba16])

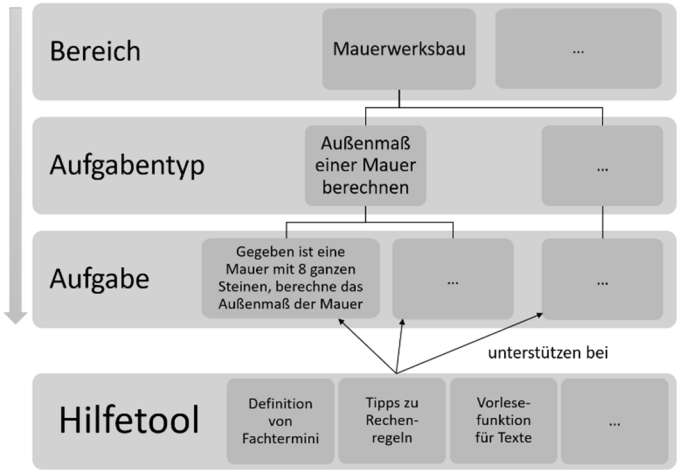
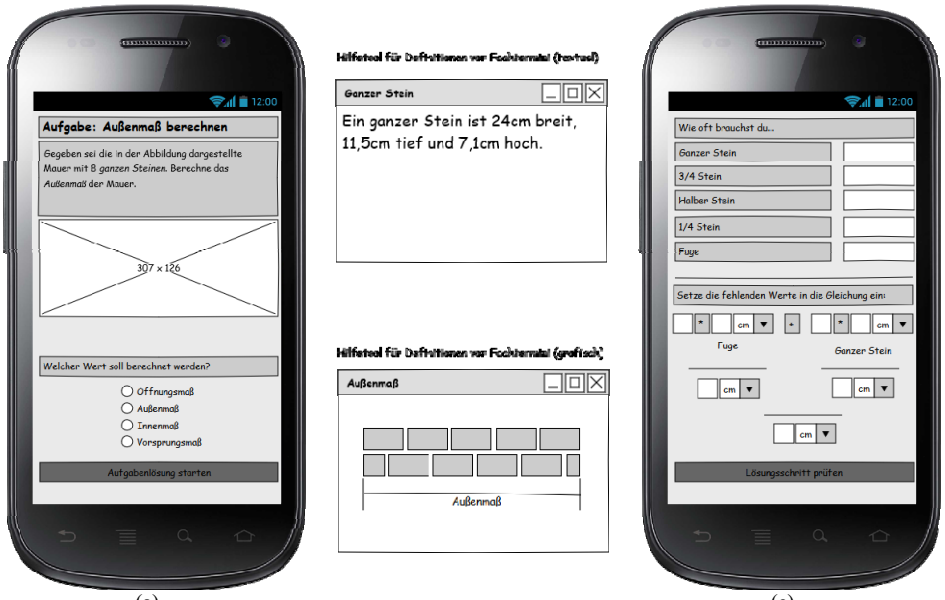



Abb. 1: Zusammenhang von "Bereich", "Aufgabentyp", "Aufgabe" und "Hilfetool" im gesamt App-Konzept

Am Beispiel eines Aufgabentyps aus dem Lehrgang zum Bereich Mauerwerksbau wird im Folgenden das Konzept für konkrete Aufgaben und die Einbindung von unterstützenden Hilfetools innerhalb dieser Aufgaben vorgestellt. Zur besseren Verständlichkeit ist in Abbildung 2 ein erster Prototyp-Entwurf für Aufgabenstellung, Hilfetool und Aufgabenlösung gegeben. Abbildung 2 (a) stellt die Aufgabenstellung dar. Je nach Aufgabentyp kann die Aufgabenstellung einen Aufgabentext, Grafiken oder andere multimediale Inhalte enthalten. Bei der späteren Umsetzung von Aufgabenstellungen ist auf eine Berücksichtigung von Richtlinien zur barrierefreien Entwicklung zu achten. Hierbei beziehen wir uns vor allem auf die Web Content Accessibility Guidelines 2.1, kurz WCAG 2.1 [W3C18]. Die Weiterentwicklung der WCAG 2.0 [W3C08] setzt hierbei einen erweiterten Fokus auf Aspekte, die für Personen mit kognitiven Einschränkungen oder Lerneinschränkungen, geringer Sehkraft oder Einschränkungen in der Nutzung mobiler Geräte, von Bedeutung sind [W3C18]. Aus diesen Gründen bieten diese Guidelines sowohl unter Berücksichtigung der Zielgruppe als auch aus technischer Perspektive eine passende Basis. So müssen beispielsweise für Grafiken erklärende Texte angegeben werden oder sollten für Texte zusätzliche Audiospuren verfügbar sein. Auf diese Art und Weise können mögliche Schwierigkeiten von Auszubildenden beim Lesen von Aufgabenstellungen oder dem Verstehen von Abbildungen reduziert werden. Wie in Kapitel zwei diskutiert führen auch Fachbegriffe dazu, dass Auszubildende die Aufgabenstellung nicht vollständig kognitiv durchdringen. An diesem Punkt kann eines der Hilfetools ansetzen. Von Aufgaben-Autoren können Fachbegriffen entsprechende Definitionen sowohl in textueller als auch in grafischer Form zugeordnet werden. Diese können dann von den Auszubildenden durch einfaches Antippen des Fachbegriffs in der Aufgabenstellung geöffnet werden. Abbildung 2 (b) zeigt einen Prototyp für dieses Hilfetool.




(a) (b) (c)

Abb. 2: (a) Beispiel für eine Aufgabenstellung, (b) Beispiel für das Hilfetool für die Definition von Fachbegriffen in textueller und grafischer Form, (c) Beispiel für die Lösungsansicht einer Aufgabe

Um sicherzustellen, dass die Auszubildenden die Aufgabenstellung erfasst haben, soll eine kurze Abfrage der zu lösenden Aufgabe korrekt beantwortet werden. Bei einer Berechnungsaufgabe könnte an dieser Stelle danach gefragt werden, welcher Wert im Anschluss berechnet werden soll (siehe Abbildung 2 (a)). Bei richtiger Beantwortung gelangen die Auszubildenden in den Lösungsscreen. Ein Beispiel hierfür ist in Abbildung 2 (c) zu sehen. Während diese prototypische Darstellung alle Schritte direkt anzeigt, ist bei der Umsetzung darauf zu achten, dass diese schrittweise freigeschaltet werden. Hierdurch sollen die Auszubildenden dazu angeleitet werden die Aufgaben systematisch und schrittweise zu bearbeiten. Weiterhin ermöglicht dieses Vorgehen, frühzeitig Feedback bei möglichen Problemen zu geben. Im Beispiel der in Abbildung 2 gezeigten Aufgabe sollen die Auszubildenden im ersten Schritt die Anzahl der benötigten Steine und Fugen eintragen. Durch frühzeitige Erkennung von Schwierigkeiten, die z. B. durch wiederholt fehlerhafter Angabe der Anzahl von benötigten Fugen zu identifizieren wären, könnten so gezielte und individuelle Hilfestellungen gegeben werden. Hierzu stehen zum einen die implementierten Hilfetools zur Verfügung, die Unterstützung bei der Bearbeitung der Aufgaben bieten. Zum anderen können, je nach erkanntem Fehler und Fehlerart, Tipps und Hinweise bereitgestellt werden.

## 4 Technische Rahmenbedingungen und geplanter Einsatz

Zur Ermittlung der technischen Rahmenbedingungen wurde eine Befragung von 92 Auszubildenden des ersten Lehrjahres am Ausbildungszentrum Kerpen durchgeführt. Alle Befragten besitzen ein Smartphone. Ca. 2/3 dieser Geräte laufen mit dem Betriebssystem Android und 1/3 mit iOS. Daher wird für die Umsetzung der mobilen Applikation ein Cross-Plattform-Ansatz unter Nutzung von Cordova[5] gewählt. Auf diese Weise können mit der gleichen Codebasis, basierend auf HTML, CSS und JavaScript, Applikationen sowohl für Android als auch für iOS Geräte generiert werden. Dies ermöglicht es zum einen allen Auszubildenden die Applikation und zum anderen diese mit ihrem eigenen Gerät zu nutzen. Auf diese Weise können Probleme, welche evtl. bei der Nutzung fremder Geräte auftauchen könnten, ausgeschlossen werden.

Die hier konzipierten Aufgaben und Hilfetools sollen im Rahmen eines iterativen Software-Entwicklungsprozesses mit Teilnehmerinnen und Teilnehmern der Zielgruppe erprobt und weiterentwickelt werden. Die Aufgaben, in denen die Hilfetools eingesetzt werden, orientieren sich dabei an den Themen der Grundlehrgänge im ersten Lehrjahr am Ausbildungszentrum Kerpen. Im Verlauf des Projekts wird so eine Toolbox entstehen, die unterschiedliche Hilfetools, wie das hier vorgestellte „Definitionstool“ beinhalten wird, welche Auszubildende mit unterschiedlichen Lernbarrieren bei der Lösung der gestellten Aufgaben unterstützen sollen. Geplant ist ein anschließender Einsatz der entstehenden Aufgaben und Hilfetools in den Grundlehrgängen der Ausbildung. Dabei gilt es den Einsatz der Toolbox exemplarisch in der Bauindustrie zu erproben. Bei der technischen Umsetzung liegt ein besonderer Fokus auf einer modularen Entwicklung der Toolbox, die eine Nutzung der Hilfetools in andere Applikationen und Domänen ermöglichen soll.

Literaturverzeichnis

[Ba16] Batran, B. et.al: Lernfeld Bautechnik Grundstufe. 13. Überarbeitete Auflage, Handwerk und Technik GmbH, Hamburg, 2016.

[Fi06] Fischedick, M.: Handelnd lernen in der Bauwirtschaft: Projektaufgaben Maurer, -in. Unterlagen für Ausbilder, Christiani, 2006.

[KMK18] Sekretariat der Kultusministerkonferenz: Sonderpädagogische Förderung in Schulen 2007 bis 2016. Statistische Veröffentlichung der Kultusministerkonferenz, Dokumentation Nr. 12, Berlin, 2018.

[W3C18] W3C, World Wide Web Consortium: Web Content Accessibility Guidelines (WCAG) 2.1 (2018), https://www.w3.org/TR/WCAG21/, Stand: 10.03.2019

[W3C08] W3C, World Wide Web Consortium: Web Content Accessibility Guidelines (WCAG) 2.1 (2008), https://www.w3.org/TR/WCAG20/, Stand: 10.03.2019

[5] https://cordova.apache.org/

# Was benötigen Hochschulen, um digitale Barrierefreiheit als Teilhabemöglichkeit umzusetzen?

Linda Rustemeier, Sanja Grimminger und Sarah Voß-Nakkour[1]

**Abstract:** Eine Teilhabe an Bildung und Wissenschaft schließt auch Menschen mit Beeinträchtigungen oder Behinderungen an Hochschulen ein. Es gibt umfassende rechtliche Vorgaben und Empfehlungen, doch leider noch keine einheitliche Barrierefreiheit an Hochschulen, insbesondere in den Bereichen Websysteme und deren Inhalte (Dokumente, Videos, Bilder, Navigationsmöglichkeiten). Dieses Positionspapier betrachtet zum einen die Vorgaben der BITV 2.0, WCAG 2.1 sowie der EU-Richtlinien von 2016 und richtet dahingehend zum anderen den Fokus auf Herausforderungen und Empfehlungen. Auch das diesjährige DeLFI-Motto ist passenderweise "Teilhabe an Bildung und Wissenschaft". Dabei ergeben sich einige Grundfragen, die vor der Konzeption jedes (inklusiven) Bildungsangebots stehen: Welche Bedarfe und Herausforderungen werden bezüglich barrierefreier Dokumente, Webseiten und deren Inhalte an inklusive Hochschulen gerichtet? Was brauchen Lehrende und Studierende mit Beeinträchtigungen, um zu partizipieren? Hilft eine Einheitlichkeit an digitaler Infrastruktur, wie barrierefreie Systeme (z.B. Lernplattformen), eine inklusive Hochschule zu werden?

**Keywords:** Digitale Barrierefreiheit, Hochschule für alle, Digitale barrierefreie inklusive Hochschule, Inklusion, Studieren bzw. Lehren mit Beeinträchtigung oder Behinderung, Diversity, Teilhabe, BITV, WCAG, EU-Richtlinie Barrierefreiheit.

## 1 Einleitung: Bedarfe und Richtlinien

Laut der "Best 2 Studie - beeinträchtigt studieren" des *Studentenwerks* haben rund 11 % der fast 2,8 Mio. Studierenden in Deutschland eine studienrelevante Beeinträchtigung [PHZB18]. Die Bedarfe der Studierenden und der Mitarbeiter*innen müssen bei einer "Hochschule für alle" auch durch eine digitale Barrierefreiheit beachtet werden. Es gibt umfassende rechtliche Vorgaben und Empfehlungen, doch leider noch keine einheitliche Barrierefreiheit an hochschulweiten digitalen Systemen, deren Inhalten wie Dokumenten, Videos, Bilder sowie Einstellungs- und Navigationsmöglichkeiten [ST17]. Eine Hochschule für alle ist eine inklusive Hochschule, alle Hochschulangehörigen sollen gleichberechtigt und ohne Diskriminierung mit anderen an der Hochschulbildung teilnehmen können [Kl16]. Bereits 2002 gab das Deutsche Studentenwerk aufgrund des Bundesverwaltungserlasses der Barrierefreien-Informationstechnik-Verordnung (BITV) eine Anleitung zur Gestaltung von barrierefreien Webseiten heraus. Die aktuelle Lage zeigt, dass z.B. viele Internetauftritte von deutschen Hochschulen immer noch nicht ausreichend barrierefrei sind und die Anpassungen im Nachhinein eine mühevolle

[1] **studium**digitale – Zentrale eLearning-Einrichtung, Goethe Universität Frankfurt, Medienproduktion, Varrentrappstraße 40-42, 60486 Frankfurt, rustemeier@sd.uni-frankfurt.de, grimminger@sd.uni-frankfurt.de, voss@sd.uni-frankfurt.de.

doi:10.18420/delfi2019_321

Aufgabe ist: Mangelnde digitale Barrierefreiheit ist kein Universitäts- oder Einzelproblem, sondern ein Problem im gesamten Web [Gr17]. Universitäten sind als öffentliche Einrichtungen durch die EU-Richtlinien dazu verpflichtet digitale Barrierefreiheit umzusetzen (für öffentliche Internetauftritte und -angebote, Intranetauftritte und –angebote, grafische Programmoberflächen, Apps und sonstige Anwendungen für mobile Endgeräte) [BMJV16].[2] Im Wesentlichen sind der weltweite Standard für barrierefreies Webdesign die *Web Content Accessibility Guidelines*, *WCAG 2.1*. Daraus entstanden auch die deutsche *BITV 2.0* [BMJV16][3] sowie die EU-Richtlinien [EU16]. Das Ziel der *BITV 2.0* ist es, Webseiten und andere grafische Oberflächen technisch so zu gestalten, dass Menschen mit Beeinträchtigungen diese digitalen Präsenzen nutzen können. Dies ist essentiell für Menschen mit körperlichen Einschränkungen, die die (Computer-)Peripheriegeräte nicht nutzen können. Das betrifft z.B. körperbehinderte Menschen, Menschen mit Sehbeeinträchtigungen, die einen starken Kontrast benötigen oder eine Lupe verwenden oder Blinde, die Screenreader einsetzen.[4] Aus den Bedarfen der Menschen bzw. den Anforderungen an die Systeme der Informationsverarbeitungen ergeben sich Herausforderungen, die im nächsten Abschnitt näher erläutert werden.

## 2 Herausforderungen

Wer Inklusion an den Hochschulen implementieren und dazu sensibilisieren will, der muss an den Hochschulen die digitale Barrierefreiheit als Priorität und mit Ressourcen umsetzen. Dabei sind die vier Prinzipien bzw. Richtlinien *wahrnehmbar, bedienbar, verständlich und robust d*er *WCAG 2.1* [W3C18] zu befolgen. Neben Richtlinien sind es aber insbesondere auch Leitfäden, die es Hochschulmitarbeitenden ermöglichen, in das Thema einzuführen und bei der Umsetzung anzuleiten. Für Lehrende stehen an vereinzelten Hochschulen Leitfäden oder Weiterbildungen zur barrierefreien Lehre im Studium zur Verfügung. Bisher fehlen in diesen Publikationen jedoch Beispiele und konkrete Anleitungen, die es Lehrenden und anderen Akteur*innen ermöglichen einen praxisnahen Eindruck von der Wirkung und der Gestaltung von barrierefreien (Lehr-) Materialien zu erhalten. Darüber hinaus müssen die technische Infrastruktur und Systeme (LMS, Webauftritte etc.) der Hochschulen sowie die Gesamtprozesse evaluiert werden, um deren Erfolge bzw. fehlende Elemente zu ermitteln. Mangelnde Medienkompetenzen im Bereich der digitalen Barrierefreiheit sollen durch kurze Ratgeber (z.B. Tutorials), Best

[2] Die Gestaltung der Angebote der Informationstechnik sollen Menschen mit Behinderungen im Sinne des § 3 des Behindertengleichstellungsgesetzes (BGG)- denen ohne die Erfüllung zusätzlicher Bedingungen die Nutzung der Informationstechnik nur eingeschränkt möglich ist- den Zugang eröffnen.

[3] Zur Erläuterung der BITV 2.0 hat das Bundesministerium für Arbeit und Soziales das Online-Portal BITV-Lotse entwickeln lassen: www.bitv-lotse.de.

[4] Der §4 des BGG besagt, dass Barrierefreiheit technische Gebrauchsgegenstände, Systeme der Informationsverarbeitung, akustische und visuelle Informationsquellen und Kommunikationseinrichtungen sowie andere gestaltete Lebensbereiche betrifft und diese für Menschen mit Behinderungen in der allgemein üblichen Weise, ohne besondere Erschwernis und grundsätzlich ohne fremde Hilfe auffindbar, zugänglich und auch mit Hilfsmitteln nutzbar gemacht werden sollen [Du18].

Practice oder Umsetzungsdienste kompensiert werden. Dazu braucht es zugängliche, praxisnahe Beispiele und Berührungspunkte, die ihre Notwendigkeit unterstreichen bspw. durch ein ergänzendes „*Leitbild inklusiv- digitale Lehre*“. Auch Beratungsangebote können Türen öffnen und eine Reihe an hilfreichen Tools stehen kostenlos zur Verfügung. Eine Zusammenfassung von Überprüfungstools der Barrierefreiheit bietet u. a. die *Web Accessibility Evaluation Tools List* [W3C18].[5] So sehr diese Tools helfen, bleibt die Frage, ob alle Hochschulmitarbeiter*innen diese Tools eigenständig beherrschen müssen. Doch es stellt sich die Frage, ob Umsetzungsdienste eine Alternative wären, die laut einer Umfrage des Studentenwerks an (mindestens) 18 von 99 deutschlandweiten teilnehmenden Hochschulen angeboten werden [ST17].[6] Aus den in diesem Kapitel beschriebenen Herausforderungen ergeben sich folgende konkrete Handlungsempfehlungen.

## 3 Handlungsempfehlungen

Aus den Richtlinien der *WCAG* und *BITV,* den Bedarfen und den Herausforderungen der Hochschulen ergibt sich im Folgenden eine erweiterbare Zusammenstellung, die in weiteren Arbeiten der Autor*innen noch evaluiert wird. Um auf die zu Beginn gestellte Leitfrage *„Was benötigen Hochschulen, um digitale Barrierefreiheit als Teilhabemöglichkeit umzusetzen?“* zurückzukommen, wurden folgende konkrete Handlungsempfehlungen ausgearbeitet:

**Deutschlandweite Maßnahmen:** Wie im ersten Kapitel beschrieben wurde, gibt es bereits umfassende Richtlinien und gesetzliche Vorgaben. Um diese erfüllen zu können, müssen Politik und Hochschulen zusammenarbeiten, denn Hochschulen benötigen Ressourcen, mit denen Sie Barrierefreiheit umsetzen können. Da die Umsetzung nicht nur Sensibilisierung, sondern auch Zeit und Geld in Anspruch nimmt, ist dies eine wesentliche Voraussetzung. Eine Überprüfung der Einhaltung zur Umsetzung von Barrierefreiheit soll zukünftig durch landesweiten Überwachungs- und Prüfstellen erfolgen (EU-Richtlinien). Die Zusammenarbeit auf oberster Ebene und ein Landesbudget für digitale Barrierefreiheit bilden das Fundament für die weiteren Schritte. Des Weiteren ist zu diskutieren inwieweit eine gewisse Einheitlichkeit von Systemen (z.B. durch Empfehlungen) zu implementieren ist, ohne die Bedarfe der Akteur*innen zu vernachlässigen.

[5] Beispiele sind: Webseitenüberprüfungen, PDF Accessibility Checker, Farb-Kontrast-Checker, Screenreader optischen Zeichenerkennung, barrierefrei-bedienbare Audio- und Videoplayer, Tools für Gebärdenvideos, Audiodeskriptionen, Alternativtextsoftware, Untertitel, Bildvergrößerungssoftware (Lupe).

[6] Umfrage zu Umsetzungsdiensten mit Beauftragten und Berater*innen für Studierende mit Behinderungen vom April 2017 des IBS des Deutschen Studentenwerks: Unter 308 Hochschulen (staatliche, private und kirchliche Hochschulen mit staatlicher Anerkennung, Verwaltungsfachhochschulen) nahmen 99 Hochschulen teil, darunter 46 Fachhochschulen, 40 Universitäten, 14 Kunst- und Musikhochschulen sowie ein Universitätsklinikum (Angaben zu 19 Umsetzungsdiensten).

**Hochschulübergreifende Maßnahmen:** Um das Thema Barrierefreiheit planbar und überprüfbar zu machen, muss es eine umfassende *Analyse und Bewertung des Status Quo* sowie eine *Bedarfsanalyse* geben. Dabei sollen Bedarfe identifiziert und kategorisiert werden, die in regelmäßigen Abständen evaluiert werden. Anhand eines Konsenses über die Bedarfe wird es erst möglich strategisch vorzugehen. Die *Kommunikation zwischen verschiedenen Akteur*innen* ist dabei essentiell, da die Prozesse bzw. die Prozessentwicklungen nur dann zufriedenstellend sein können, wenn nicht nur Expert*innen (Landes- und Bundesakteure digitaler Barrierefreiheit), sondern auch Betroffene eingebunden werden. Erst ein starkes Netzwerk macht es möglich, einen umfassenden *Erfahrungsaustausch* zu gewährleisten, um daraus z.B. Inklusionsposten und Budgetfragen zu klären sowie gemeinsam politische Forderungen aufzustellen. Innovative Schulungsstrategien können helfen, (digitale) Möglichkeiten und Spielräume zu entwerfen. Auch könnte ein „Forum digitale Barrierefreiheit" einen fruchtbaren Onlineaustausch an Erfahrungswerten generieren oder geeignete Lösungen finden.

**Hochschulinterne Maßnahmen:** Für digitale Barrierefreiheit ist die Förderung der Maßnahmen durch die Hochschulleitung unabdingbar, da nur so das Thema in alle Bereiche mit hineingetragen werden kann und es nicht nur bei Vorhaben bleibt, sondern diese auch umgesetzt werden. Nur wenn die Sensibilisierung für das Thema auf allen Ebenen vorhanden ist, führt dies zu einer inklusiven Hochschule. Erstrebenswert wäre der Einsatz einer/s Experten/in für digitalen Barrierefreiheit im Präsidium bzw. die/der als Koordinator*in fungiert und alle strategischen Entscheidung über neue digitale Projekte prüft und in Planungs- und Umsetzungsprozessen eingebunden ist. Bei der Beschaffung von Systemen von Drittanbietern, muss digitale Barrierefreiheit verpflichtend sein. Denn ist das verwendete CMS nicht zugänglich, bzw. lassen sich damit keine barrierefreien Inhalte erstellen, können auch alle Workshops wenig beisteuern. Damit alle Akteur*innen an Hochschulen die digitale Barrierefreiheit umsetzen können, ist es wichtig, neben Sensibilisierungsmaßnahmen, zum einen Schulungen und Weiterbildung zu entwickeln (sowohl für Entwickler*innen von Systemen als auch für Autor*innen von Inhalten) und zum andern praxisnahe Leitfäden, Checklisten und konkrete Anleitungen zu erstellen. Daneben soll individuelle Beratung und Unterstützung durch z.B. Umsetzungsdienste, eLearning-Einrichtungen, oder Serviceleistungen der Hochschulrechenzentren, eine bedarfsorientierte Hilfe gewährleisten: in Form von Support, Tutorials, Best-Practice Beispielen, Dokumentation etc.
Studierende sollen nach dem Credo *„Nicht über uns - ohne uns"* aktiv in Planungen durch das Studien-Service-Center sowie Beratungen in Kooperation mit dem Lehrenden und Prüfungsämtern einbezogen werden. Die individuellen Bedarfe von Studierenden können weiterhin durch *Inklusionspläne* nach angloamerikanischen Vorbild von Anfang an identifiziert und die Beteiligten können auf diese Weise besser unterstützt werden.
Ein letzter Punkt, der zum Nachdenken über eine mögliche Umsetzung einlädt: Es stellt sich die grundsätzliche Frage, ob eine Vereinheitlichung der verwendeten Systeme sinnvoll sein kann. Diese ist wohl ein Wunschtraum von Betroffenen, solange es keine passenden *Add-ons* für die Bedarfe der Akteur*innen gibt. Wenn es einheitliche,

vorgegebene Hochschulsysteme (wie Lernmanagementsysteme[7] wie Moodle, Mailsysteme und CMS Cloud-Lösungen etc.) geben würde (Freiheit der Systemwahl- und Marktführer- oder Open Source-Probleme eingeschlossen), würde das viele Prozesse erleichtern, da alle Akteur*innen ihre Ressourcen bündeln könnten.
Alle Maßnahmen auf allen Ebenen sollten regelmäßig evaluiert werden und anhand von Empfehlungen zur Weiterentwicklung beitragen. Diese sollten dann wieder deutschlandweit bearbeitet werden. Den Abschluss des Positionspapiers bilden das Fazit und der Ausblick zum Thema digitale Barrierefreiheit an Hochschulen.

## 4 Fazit und Ausblick

Weder die Mitarbeiter*innen noch die einzelnen Hochschulen sollten zeitintensive Einzelkämpfe ausrichten, sondern durch landes- oder bundesweite Stellen unterstützt werden, Barrierefreiheit umzusetzen. Ein regelmäßiger Austauschprozess durch Publikationen und zentrale Weiterbildungs-, Beratungs- und Infostellen sowie Kontrollinstanzen sind wichtig für (digital- inklusive) Hochschulen. Ebenso sind neben der Aneignung eines Know-Hows und einem Aufwand an Vernetzungsaktivitäten, auch ein Budget bzw. Verantwortlichkeiten noch nicht einheitlich geregelt – das muss sich ändern. Auch wird der Vorteil des Internets in Bezug auf Vernetzung noch unzureichend genutzt: Bei aufkommenden technischen Problemen könnten durch hochschulweite Crowd-Lösungen wie z.B. ein zentrales Forum wiederkehrende Probleme einfach und schnell gelöst werden. Neben den barrierefreien IT-Richtlinien, *wahrnehmbar, bedienbar, verständlich* und *robust*, muss Barrierefreiheit politisch auf allen Ebenen mitgedacht werden. Die Prozesse müssen, unabhängig von Studierendenzahlen und Kosten, transparent, priorisierend und verständlich an die Universitätsangehörigen herangetragen werden. Die Kombination aus Sensibilisierung, Weiterbildung, Austausch, Zusammenarbeit, Netzwerken und hochschulinterne Überprüfungs- bzw. hochschulexterne Kontrollinstitutionen könnte ein geeignetes Mittel sein, um digitale Barrierefreiheit wirksam an den Hochschulen umzusetzen. Das würde die Hochschulen befähigen, ihrer staatlichen Verpflichtung eine *Hochschule für alle* zu sein, nachzukommen und es Menschen mit Beeinträchtigungen durch digitale Barrierefreiheit ermöglichen, eine inklusive Teilhabemöglichkeit auszuleben. Dazu braucht es Handlungskompetenz, Mut, kreative und flexible Experimentierfreude und Bewusstsein dafür, digitale Barrierefreiheit ernst zu nehmen, Budgets für digitale Barrierefreiheit zu vergeben und sie als Selbstverständlichkeit von Anfang an, statt nachträglich, einzubinden. Diese notwendigen Schritte zu einer *Hochschule für alle* durch digitale Barrierefreiheit sind weder aufzuschieben, noch ist an diesem Grundrecht zu sparen.

[7] Für Barrierefreiheit und Anpassbarkeit entwickelte Lernplattformen sind z.B. A-Tutor oder die BFW online Plattform für Blinde. Weitere LMS wie ILIAS, Moodle, OPAL oder Sakai sind für Barrierefreiheit weiterentwickelt worden. Dabei sind einzelne Module unterschiedlich gut durch assistive Technologie nutzbar [SbR18].

## Literaturverzeichnis

[BFS18] Bundesfachstelle Barrierefreiheit: https://www.bundesfachstelle-barrierefreiheit.de/DE/Praxishilfen/Informationstechnik/informationstechnik_node.html, Stand: 21.02.2019.

[BMJV16] BIT-V (Webseiten und Autorensysteme): https://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html, Stand: 05.02.2019.

[BMJV18] Behindertengleichstellungsgesetz (BGG): https://www.gesetze-im-internet.de/bgg/BGG.pdf, Stand: 05.02.2019.

[Du18] Dusel, J.: Behindertenbeauftragter der Bundesregierung: https://www.behindertenbeauftragter.de/DE/Themen/Barrierefreiheit/WasistBarrierefreiheit/WasistBarrierefreiheit_node.html, Stand: 03.03.2019.

[EU16] EUR-LEX: RICHTLINIE (EU) 2016/2102 DES EUROPÄISCHEN PARLAMENTS UND DES RATES, über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen, https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32016L2102, Stand: 03.03.2019.

[Gr17] Gröber, P.: Barrierefreie Webseite für Blinde? In: Eibl, M. & Gaedke, M. (Hrsg.), INFORMATIK 2017. Gesellschaft für Informatik, Bonn, S. 331-343, 2017.

[Kl16] Klein, U.: Inklusive Hochschule: neue Perspektiven für Praxis und Forschung. Beltz Juventa, Weinheim, Basel, S.7, 2016.

[PHZB18] Poskowsky, Jonas; Heißenberg, Sonja; Zaussinger, Sarah; Brenner, Julia: Beeinträchtigt studieren-best2: Datenerhebung zur Situation Studierender mit Behinderung und chronischer Krankheit 2016/17. Deutsches Studentenwerk (DSW), Berlin, 2018. online unter: https://www.studentenwerke.de/de/content/barrierefreie-hochschule-%E2%80%93-gesetzliche-grundlagen-richtlinien-und-empfehlungen, Stand: 03.03.2019.

[ST02] Deutsches Studentenwerk: https://www.studentenwerke.de/sites/default/files/Barrierefreies_Internet.pdf, Stand: 03.03.2019.

[ST17] Deutsches Studentenwerk: https://www.studentenwerke.de/sites/default/files/ibs_umfrage_2017_umsetzungsdienste_hochschulen_ergebnisse.pdf, S. 2-3, Stand: 03.03.2019.

[SbR18] e-teaching.org, Stiftung Medien in der Bildung (SbR): https://www.e-teaching.org/didaktik/konzeption/barrierefreiheit, Stand: 13.02.2019.

[W3C18] WC3: WCAG2.1: https://www.w3.org/TR/WCAG21/ und Tools https://www.w3.org/WAI/ER/tools/, Stand: 04.03.2019.

# Semantic Textual Similarity von textuellen Lernmaterialien

Moritz C. Rieger[1] und Niels Seidel[2]

**Abstract:** Diese Arbeit zeigt eine Methode, mit welcher semantische Ähnlichkeiten zwischen Lehrveranstaltungen anhand textueller Lernmaterialien erkannt und visualisiert werden können. Dadurch soll es Studierenden und Lehrenden erleichtert werden, Relationen zwischen den Lehrveranstaltungen zu entdecken. Um dieses Ziel zu erreichen, wurde exemplarisch ein vollständiger Satz von 572 Studienbriefen der Fakultät für Mathematik und Informatik der FernUniversität in Hagen als *Document Embeddings* vektorisiert und anschließend über die Kosinusähnlichkeit der Vektoren verglichen. Der gesamte Prozess kommt dabei ohne gelabelte Daten aus. Das Ergebnis der Analyse weist eine hohe Pearson's Korrelation zu einem von Domänenexperten ermittelten Goldstandard auf. Die Anwendungsmöglichkeiten dieser Ergebnisse werden am Beispiel eines Empfehlungssystem, eines Learning Dashboard für Lehrende und in einer Komponente einer adaptiven Lernumgebung demonstriert.

**Keywords:** NLP, Semantic Textual Similarity, Document Embedding, Learning Analytics

## 1. Einleitung

Vor jedem Semester stehen Studierende vor der Frage, welche Lehrveranstaltung sie belegen sollen. Für die Erreichung des Studienziels innerhalb eines Studiengangs finden sich in den Prüfungsordnungen Angaben zu den fakultativen und obligatorischen Modulen und Veranstaltungen. Studienpläne der Studienberatung flankieren diesen Rahmen mit Empfehlungen zur Anzahl, Reihenfolge und Auswahl von Lehrveranstaltungen für die einzelnen Studiensemester. Letztendliches ergeben sich aus den Terminen der Veranstaltungen weitere organisatorische Anforderungen, mit denen der individuelle Stundenplan in Einklang gebracht werden muss. Trotz dieser organisatorischen Beschränkungen eröffnet die innere Autonomie der Hochschulen viele Wahlmöglichkeiten, um Veranstaltungen gemäß inhaltlicher Kriterien und Interessen auszuwählen. Für die Entscheidungsfindung stehen jedoch in der Regel nur Modulhandbücher und Veranstaltungswebseiten zur Verfügung. Vorab als Lehrbuch oder im Sinne von OER veröffentlichte Lernmaterialien sind die Ausnahme. Die Menge an Informationen ist in beiden Fällen jedoch nur schwer zu bewältigen. Die lineare Aufbereitung der oftmals über einhundert Seiten umfassenden Modulhandbücher erschwert das Erkennen von inhaltlich ähnlichen und aufeinander aufbauenden Lehrveranstaltungen. Die knappen Beschreibungen der Module bilden zudem nur einen Bruchteil der Lerninhalte ab. Veranstaltungen, die nicht dem Studiengang zugeordnet sind, erscheinen selbstverständlich nicht im

[1] FernUniversität in Hagen, Lehrgebiet Kooperative Systeme, Universitätsstraße 1, 58084 Hagen, moritz.rieger@studium.fernuni-hagen.de
[2] FernUniversität in Hagen, Lehrgebiet Kooperative Systeme, Universitätsstraße 1, 58084 Hagen, niels.seidel@fernuni-hagen.de

doi:10.18420/delfi2019_360

Modulhandbuch. Viele Studierende holen sich deshalb bei Freunden und Kommilitonen Rat oder orientieren sich an Empfehlungen von Lehrenden. Studieninteressierte und Studienanfänger verfügen jedoch noch nicht über diese Kontakte. Besonders deutlich wird diese Herausforderung am Beispiel der FernUniversität in Hagen. Mit über 74.000 eingeschriebenen Studierenden und einem Kursangebot von über 1.600 Kursen ist sie die größte Universität Deutschlands. Allein auf die Fakultät Mathematik und Informatik entfallen 134 Lehrveranstaltungen, davon 94 Kurse und 40 Seminare bzw. Praktika. Für Studierende an der Fakultät stellt die Wahl aus dieser großen Menge an Lehrveranstaltungen eine besondere Herausforderung dar, da in der Regel nicht, wie an Präsenzuniversitäten möglich, auf das Erfahrungswissen von Kommilitonen zurückgegriffen werden kann. Darüber hinaus sind die Autoren beziehungsweise Betreuenden den Studierenden meist nicht persönlich bekannt, sodass Kontakte zu Lehrenden kaum die Entscheidung erleichtern können. Für die Entscheidungsfindung stehen den Studierenden die kurzen Beschreibungen von Veranstaltungsinhalten und Lernzielen in den Modulhandbüchern (ca. 100–200 Wörter), sowie kurze Leseproben eines Kapitels des Skriptes zur Verfügung. Für Hochschulen mit sehr vielen Lehrveranstaltungen und sehr großen Wahlmöglichkeiten ist die Planung des Studiums somit zeitaufwändig und komplex.

Vor einer vergleichbaren Hürde stehen auch Lehrende, die bei der Planung und Erstellung von Lernmaterialien einerseits Redundanzen zu anderen Kursen vermeiden und andererseits an Vorwissen anknüpfen oder selbiges für andere Veranstaltungen entwickeln möchten. Angesichts der großen Anzahl an Kursen, wissen die Betreffenden jedoch nicht immer genau, was die Kollegen in ihren Kursen im Detail lehren. In Konsequenz bleiben inhaltliche Überschneidungen und auch Lücken unentdeckt und Potentiale für eine Zusammenarbeit auf dem Gebiet der Lehre werden nicht erkannt.

In diesem Beitrag wird ein Verfahren zur Analyse der semantischen Ähnlichkeit von Lehrveranstaltungen anhand von textbasierten Lernmaterialien vorgestellt. In den folgenden Abschnitten wird zunächst der Stand der Forschung hinsichtlich Verfahren zur Bestimmung der semantischen Ähnlichkeit von Texten sowie auf Kurswahl-Empfehlungssysteme dargestellt. Im Anschluss wird das hier eingesetzte Verfahren zur Analyse semantischer Ähnlichkeiten am Beispiels eines Korpus von 94 Lehrveranstaltungen der FernUniversität in Hagen vorgestellt. Die aus dem Verfahren resultierenden Daten werden in Abschnitt 4 in einem Empfehlungssystem, einem Learning Dashboard für Lehrende und in einer Komponente einer adaptiven Lernumgebung in praktische Anwendungen überführt.

## 2. Stand der Forschung

Die Verarbeitung natürlicher Sprache mit Hilfe von Techniken des *Natural Language Processings* (NLP) konnte in den letzten Jahren enorme Fortschritte verzeichnen. Herkömmliche NLP-Methoden erzeugen aus einem Textdokument durch *Bag of Words* (BOW), frequenzbasierte Verfahren wie *Term Frequency* Inverse *Document Frequency*

(TF-IDF), *Latent Dirichlet Allocation* (LDA) usw. Vektoren und berechnen den Abstand zwischen den Vektoren [Li18]. Diese Methoden können jedoch die semantische Distanz nicht erfassen oder sind sehr rechenintensiv [Mi13] erzielen in der Regel keine guten Ergebnisse. Neuere maschinelle Lernverfahren erzielen bei der Analyse semantischer Textrepräsentationen wesentlich bessere Ergebnisse [DOL15]. Eine zentrale Herausforderung besteht dabei in der Bestimmung der semantischen Ähnlichkeit von Texten (*Semantic Textual Similarity, STS*), welche in der maschinellen Übersetzung, der semantischen Suche, der Beantwortung von Fragen und in Chatbots ihre Anwendung findet. Mit Hilfe von Entwicklungen im Bereich der *Distributed Representations*, insbesondere durch Neuronale Netze und *Word Embeddings* wie *Word2Vec* [Mi13] und *Glove* [PSM14], können semantische Eigenschaften von Wörtern in Vektoren abgebildet werden. Le und Mikolov haben mit Doc2Vec gezeigt, dass sich die dabei verwendeten Prinzipien auch auf Dokumente angewendet werden können [LM14].

Die Ähnlichkeit von umfangreichen Buchkollektionen wurde bislang nur in wenigen Arbeiten untersucht. Die von Kiros et al. vorgestellten *SkipThrough Vectors* [Ki15] trainieren ein Encoder-Decoder-Modell, das versucht, die umliegenden Sätze einer kodierten Passage zu rekonstruieren. Die Experimente basierten jedoch auf einem vergleichsweise kleinem Korpus von nur 11 Büchern [Zh15]. Spasojevic und Pocin ermittelten hingegen die semantische Ähnlichkeit auf der Ebene einzelner Seiten und ganzer Bücher für den Korpus von *Google Books*, der ca. 15 Millionen Bücher umfasst [SS11]. Die Ähnlichkeit zweier Bücher wurde aus dem *Jaccard Index* der permutierten Hash-Werte von normierten Wortgruppen (Features) ermittelt. Liu et al. verweisen jedoch darauf, dass die semantische Struktur längerer Dokumente auf diese Art nicht berücksichtigt werden kann und schlagen deshalb die Repräsentation als *Concept Interaction Graph* vor [Li18]. Aus einem Paar von Dokumenten werden Schlüsselwörter bestimmt und mittels *Community Detection* Algorithmen zu Konzepten (Knoten) zusammengefasst. Diese Knoten sind durch Kanten miteinander verbunden, die die Wechselwirkungen zwischen den Knoten auf Basis von Sätzen aus den Dokumenten wiedergeben. Trotzdem das Verfahren sehr vielversprechend erscheint, wurde es bislang nur anhand von Nachrichtenbeiträgen untersucht. Die *SemEval-2018 Task 7* [Ga17] verfolgt in Bezug auf STS beispielsweise eine ähnliche Zielstellung wie dieser Beitrag, in dem semantische Relationen aus Abstracts wissenschaftlicher Artikel gefunden werden sollen. Der zur Evaluation verwendete Goldstandard basiert auf *Named Entities* (Personen, Orte, Organisationen), welche jedoch für große Textmengen nicht mit vertretbarem Aufwand annotiert werden können.

Brackhage et al. haben Modulbeschreibungen mehrerer Hochschulen durch Experten manuell verschlagworten lassen und diese Daten zusammen mit weiteren Metadaten in einer Web-Anwendung als *Forced Layout Graphen* und Adjazenzmatrix-Heatmap veranschaulicht sowie mit Hilfe komplexer Filter durchsuchbar gemacht [Br16]. Die Verschlagwortung erwies sich jedoch als enorm zeitaufwendig und muss häufig aktualisiert werden. Baumann, Endraß und Alezard haben mit Hilfe von Studienverlaufsdaten „zum einen die Verteilung der Studierenden über die Module in einem Studiengang und zum anderen die Verteilung der Studierenden in einem Modul über verschiedene Studiengän-

ge“ visualisiert [BE15], ohne jedoch deren Nutzen für die angedachte Unterstützung bei der Wahl von Veranstaltungen zu konkretisieren. Askinadze und Conrad haben anhand von Prüfungsdaten eines Studiengangs den Studienfortschritt und -abbruch in verschiedenen Visualisierungen dargestellt [AC18].

Es gibt jedoch eine große Zahl von Anwendungen und Ansätze, um Studierenden Kurse zu empfehlen. Lin et al. entwickelten mit der *Sparse Linear* Methode topN-Emfehlungen auf Basis von Belegungsdaten bestimmter Gruppen von Studierenden [Li18]. Mit Hilfe von *K-Means* und *Apriori Assoziation Rules* haben Aher und Lobo eine Empfehlungssystem für Kurse im Lernmanagementsystem Moodle vorgestellt [AL13]. Zablith et al. stellen mehrere Empfehlungssysteme vor, die auf *Linked Data* der Open Univeryity UK[3] beruhen [ZS10]. Die Anwendung *Social Study* zum Beispiel, schlägt Lernenden anhand ihres Facbook-Profils Kurse vor, während *Linked OpenLearn* den Lernenden Medien und Kurse zu den *OER* der Fernuniversität anbietet. Die Empfehlungen beruhen auf kursbezogenen Metadaten und Verknüpfungen zu anderen Kursen sowie Medien, berücksichtigen jedoch nicht die Semantik der Kurse. D‘Aquin und Jay versuchen die fehlende Semantik mit Hilfe unterschiedlicher Linked Data Quellen (u.a. *DBpedia*) zu rekonstruieren, um damit häufig auftretende Verläufe von Kursbelegungen (*Frequent Sequences*) nachvollziehbar zu machen [LM14]. Die Analyse semantischer Ähnlichkeiten auf Grundlage der vollständigen Lernmaterialien gewährt deshalb nicht nur Einblicke in die inhaltlichen Beziehungen von Lernressourcen, sondern eröffnet auch die Möglichkeit die zeitliche Struktur und Verlaufsmuster von Kursbelegungen für die Entscheidungsfindung bei der Kurswahl verständlich zu machen.

Aus der Perspektive der Planung von Kursangeboten haben Kardan et al. mit Hilfe eines neuronalen Netzes ein Vorhersagemodell für die Anzahl an Kursbelegungen entwickelt [Ka13]. Ognjanovic et al. haben darüber hinaus die Kursbelegungen für mehrere Semester im Voraus modelliert [OGD16]. Für eine didaktisch motivierte Nutzung von Belegungsstatistiken konnten die Autoren dieses Beitrags jedoch keine Beiträge in der Literatur finden. Gleiches trifft auch für die Verwendung dieser Daten zur Modellierung von Lernenden innerhalb adaptiver oder zumindest personalisierter Lernumgebungen zu.

## 2. Ermittlung der semantischen Ähnlichkeit von Texten

In diesem Abschnitt wird ein Verfahren zur Analyse der semantischen Ähnlichkeit von Lehrveranstaltungen anhand von textbasierten Lernmaterialien am Beispiel der Studientexte der Fakultät Mathematik und Informatik der FernUniversität in Hagen vorgestellt. Dieses Verfahren setzt sich aus vier Schritten zusammen, die im Wesentlichen auf die Arbeit von Rieger [Ri18] zurückgehen. Zunächst wird ein Korpus der Kursunterlagen erstellt. Anschließend werden diese Daten vektorisiert, um im dritten Schritt die Ähnlichkeit bestimmen zu können. Abschließend erfolgt eine Evaluation mit einem Goldstandard und anderen Vergleichsgrößen.

[3] Siehe http://data.open.ac.uk/ (abgerufen am 21.07.2019).

Als Korpus bezeichnet man eine Sammlung zusammenhängender Dokumente. Zur Erstellung eines Korpus lagen Ausgangsdaten von 94 Kursen aus allen 20 Lehrgebieten der Fakultät vor. Ein Kurs besteht aus 3 bis 10 Dokumenten, den Studienbriefen. Die Studienbriefe lagen in Form von PDF-Dokumenten vor. Die PDFs unterschieden sich hinsichtlich ihres Formats (bspw. PDF/A, PDF/X), den PDF-Versionen und Werkzeugen, die für deren Erzeugung genutzt wurden. Auch die Formatierung des Satzspiegels war nicht einheitlich. Aus diesen Gründen erwies sich eine programmatische Extraktion von Kapiteln mittels regulärer Ausdrücke und PDF-Outlines als wenig zuverlässig und musste verworfen werden. Die Deckblätter sowie redundante Inhaltsverzeichnisse und Schlagwortregister innerhalb eines Kurses wurden entfernt. Die PDF-Dokumente wurden daher zunächst in Text konvertiert und mit Hilfe des NLTK [BLK09] in Sätze und Wörter unterteilt. Um Fehler mit mathematischen Formeln und Punktlinien im Inhaltsverzeichnis zu vermeiden, wurde der Text außerdem normalisiert. Der so entstandene Korpus enthält 572 Studienbriefe, bestehend aus 654.367 Sätzen mit insgesamt 9.507.770 Wörtern. Das Vokabular umfasst 179.078 verschiedene Wörter.

Zur Vektorisierung der Dokumente wurde auf *Document Embeddings*, auch genannt *Paragraph Vectors (PV)* von Le und Mikolov zurückgegriffen [LM14]. Da *Document Embeddings* auf *Word Embeddings* aufbauen, müssen diese zuerst erstellt werden. Dazu gehen die Wörter im One-Hot-Encoding in ein neuronales Netz ein. Dieses dient einer Schätzaufgabe, wobei aus dem Kontext eines Wortes, das sich am wahrscheinlichsten in diesem Kontext befindende Wort geschätzt werden soll. Das neuronale Netz wird mit Tupeln, aus dem Text trainiert. Dazu wird ein Fenster durch den gesamten Korpus geschoben und die entstehenden Tupelkombinationen innerhalb des Fensters notiert. Durch Rückführung des Schätzfehlers in die erneute Schätzung werden die Gewichte der Gewichtsmatrix W optimiert. Dies hat zur Folge, dass die Gewichte der Schätzaufgabe für semantisch nahe Wörter ähnliche Werte annehmen, da in das Training vergleichbare Tupel eingingen. Die Gewichte der Schätzaufgabe stellen die *Word Embeddings* dar.

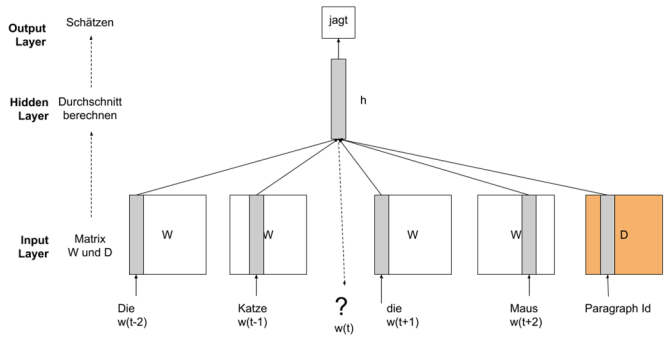



Abb. 1: *Continuous Bag-of-Words* Modell sowie die Paragraph ID (orange), die neben den Kontextwörtern bei *Document Embeddings* in die Schätzung von w(t) eingeht.

Um ganze Dokumente semantisch als Vektor abbilden zu können, wird die Idee der Word Embeddings auf ganze Texte erweitert. Dazu wird einem *Paragraph Vector,* eine

Spalte einer weiteren Gewichtsmatrix D, mit Wortvektoren kombiniert, um das nächste Wort aus einem gegebenen Kontext zu schätzen (siehe Abb. 1). Da die Wortvektoren als indirekte Folge der Schätzaufgabe die Semantik der Wörter erfassen, geschieht dies mit *Document Embeddings* auf ähnliche Weise. Man kann sich das Training der PV, wie das Training eines weiteren Wortes vorstellen. Ein PV fungiert als eine Art Speicher, welcher Informationen über fehlende Wörter im Kontext innerhalb eines Dokuments enthält. Aus diesem Grund wird dieses Modell auch *Distributed Memory Model* of PV genannt. Nachdem aufbauend auf den Word Embeddings PV trainiert wurden, um ganze Dokumente zu repräsentieren, können die PV nun als Merkmale der Dokumente weiterverarbeitet werden um semantische Ähnlichkeiten der Dokumente zu erkennen.

Bevor die Dokumente miteinander verglichen werden, wird zuerst die semantische Ähnlichkeit von Texten allgemein untersucht. Um eine Gemeinsamkeit aller Begriffe zu finden, muss Ähnlichkeit als „komplexes Netzwerk von Gemeinsamkeiten“ [Wi53] verschiedener Entitäten gedacht werden. Diese komplexe Form der Ähnlichkeit natürlicher Sprache bedeutet, dass zwei Dokumente nicht aufgrund gemeinsamer Merkmale als semantisch ähnlich angesehen werden können, sondern Ähnlichkeit aus dem Zusammenspiel vieler, direkter und indirekter Beziehungen der darin enthaltenen Wörter zu verstehen ist [Gr09]. Diesem Ähnlichkeitsbegriff wird durch das Training der *Word Embeddings* Rechnung getragen. Die Gewichtsmatrix, welche letztendlich die *Word Embeddings* enthält, ist Resultat der Verwendung der Wörter in allen Kontexten des gesamten Textkorpus und repräsentiert folglich das von Wittgenstein beschriebene komplexe Netzwerk von Gemeinsamkeiten. Zum Vergleich von *Distributed Representations* wird üblicherweise die Kosinusähnlichkeit als Metrik verwendet [WIZ15,Gr09]. Für normierte PV besteht eine lineare Beziehung zur euklidischen Distanz.

## 3. Evaluation des NLP-Systems

Da das Vektorisieren der Dokumente als PV eine *Unsupervised Machine Learning* Methode ist, gibt es keine zugrunde liegenden Testdaten gegen die das System getestet werden kann. In Anlehnung an die SemEval-Wettbewerbe wurde deshalb ein Goldstandard entwickelt, der aus einem Set an Test- und Trainingsdaten besteht. Dieser Goldstandard ließ sich jedoch nicht wie bei vielen SemEval Tasks durch *Crowdsourcing* [Ag12] erzeugen, da für die Beurteilung der semantischen Ähnlichkeit ein hohes Maß an Kompetenz in den betreffenden Wissensgebieten erforderlich ist. Aus diesem Grund wurden drei Experten, eine Autorin und zwei Autoren von Kurstexten, gebeten, einen ihrer eigenen Kurse mit einem ihrer Meinung nach ähnlichen Kurs zu vergleichen. Durch die Bewertung von thematisch bzw. inhaltlich verwandten Dokumenten ließen sich monotone Goldstandards vermeiden, die keine Übereinstimmungen aufweisen. Durch die drei Experten wurden je zwei Kurse mit jeweils 4 und 7 Studienbriefen bewertet. Jeder Bewerter hatte folglich 28 Vergleiche vorgenommen. Die Ähnlichkeit wurde auf einer kontinuierlichen Skala von 0 (nicht ähnlich) bis 100 (identisch) angegeben. Auf

eine nominale Abstufung der Skala wurde aufgrund von zu erwartenden Verständnisproblemen in Bezug auf die Wertigkeit und Äquidistanz der Skalenwerte verzichtet.

Ein Teil des Goldstandards wurde, wie in Abb. 2 dargestellt, für das Tuning der Hyperparameter verwendet. Die Hyperparameter setzten sich aus der Fenstergröße des *Continuous Bag of Words*, der Dimension des PV und der minimalen Auftrittshäufigkeit der betrachteten Wörter zusammen. Die Werte für die einzelnen Parameter beruhen auf Plausibilitätstests und sind im Rahmen der aus der Literatur bekannten Wertebereiche. Um ein Overfitting zu vermeiden, wurden die Werte je Parameter nur grob abgestuft. Der minimale mittlere quadratische Fehler konnte für eine Fenstergröße von 20, ein Dimension des PV von 140 und einem Wert von 20 als untere Schranke für die Auftrittshäufigkeit von Wörtern ermittelt werden. Auf Grundlage dieser Hyperparameter wurde ein Modell trainiert und mit dem zweiten Teil des Goldstandards getestet. Pearson's r als Maßzahl des linearen Zusammenhangs erreichte einen Wert von 0,598. Da jedoch im vorliegenden Fall ganze Dokumente, statt einzelner Sätze verglichen wurden, sind der Goldstandard und die PV unschärfer. Feine Schwankungen der Kosinusähnlichkeit werden im Goldstandard nicht abgebildet. In Anbetracht der subjektiven Einschätzung auf der von den Bewertern frei interpretierbaren kontinuierlichen Skala, ist der Wert für Pearson's r jedoch als hoch anzusehen. Zur Feststellung der monotonen Beziehung von Kosinusähnlichkeit und Goldstandard wurde Kendall's τ als Rangkorrelationskoeffizient mit einem Wert von 0,451 ermittelt. Generell erhält man kleinere Werte der Korrelation für Kendalls τ im Vergleich zu Pearson's r. Der geringe Wert ist jedoch auch auf die individuelle Definition des Ähnlichkeitsbegriffs und die individuelle Abbildung der subjektiv wahrgenommenen Ähnlichkeit auf die Skala zurückzuführen. Betrachtet man die in Abb. 3 dargestellten Bereiche großer Ähnlichkeit, so ist der Zusammenhang deutlicher zu erkennen.

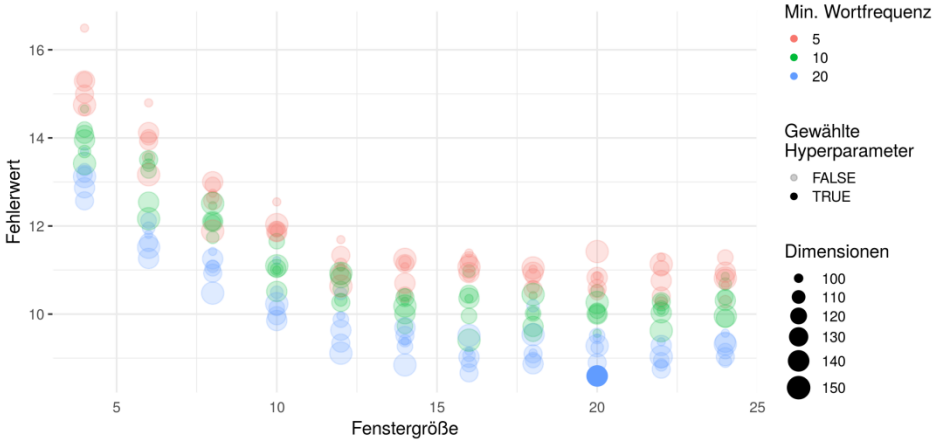


Abb. 2: Die Minimierung des mittleren quadratischen Fehlers für mehrere Konfigurationen von Hyperparametern. Blau hervorgehoben ist die beste Parameterkombination, orange dargestellt ist der Goldstandard.

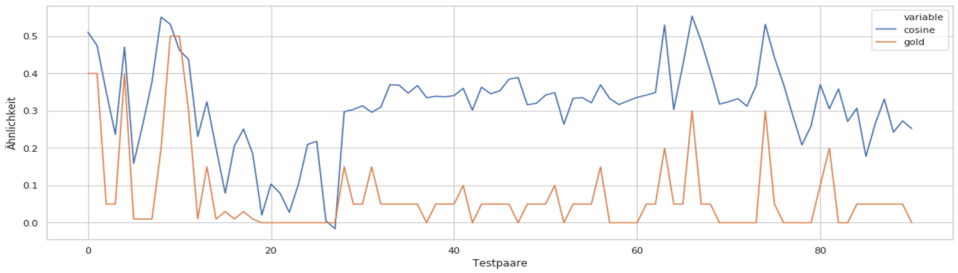



Abb. 3: Verhältnis von Goldstandard zur Kosinusähnlichkeit für die einzelnen Testpaare

Ergänzend zum Goldstandard wurde das NLP-System hinsichtlich der Plausibilität der Ergebnisse geprüft. Dazu wurden zwei Hypothesen aufgestellt: (1) Studienbriefe von konsekutiven Kursen sind sich ähnlicher, als Studienbriefe anderer Kurse. (2) Die Studienbriefe eines Kurses sind sich ähnlicher, als Studienbriefe anderer Kurse. Zur Prüfung der ersten Hypothesen wurden zunächst acht Kurse identifiziert, die angesichts der im Kurstitel enthaltenden Nummerierung eindeutig auf einander aufbauen. Die mittlere Kosinusähnlichkeit der konsekutiven Kurse liegt mit dem Wert 0,32 über dem Durchschnitt des gesamten Korpus (0,18). Hypothese 1 ist damit bestätigt. Die zweite Hypothese ließ sich bereits an den stark eingefärbten rechteckigen Artefakten entlang der Diagonalen in der Adjazenzmatrix erkennen. Die mittlere Ähnlichkeit von Studienbriefen eines Kurses beträgt 0,51 und ist somit deutlich größer als die mittlere Kosinusähnlichkeit des gesamten Korpus. Auch Hypothese 2 ist demnach bestätigt. Ein weiterer Teil der Plausibilitätsprüfung bestand u.a. darin, unerwünschte Effekte der Dokumentgröße auf die semantische Ähnlichkeit auszuschließen. Ein Zusammenhang zwischen der Differenz der Wortanzahl zweier Dokumente und ihrer Kosinusähnlichkeit besteht nicht ($r = 0{,}013$).

## 4. Anwendungen

Um die extrahierten Relationen für Anwender zugänglich zu machen, wurden mehrere interaktive Diagramme als *Data Driven Documents* [BOH11] erstellt und in eine Web Anwendung überführt. Neben den semantischen Ähnlichkeiten von Kursen und Studienbriefen konnten auch historische Daten über die von Studierenden belegten Kurse einbezogen werden. In diesem Datensatz sind alle Kursbelegungen seit der Einführung der Bachelor- und Masterstudiengänge im Jahre 2003 an der Fakultät Mathematik und Informatik je Semester und Person enthalten.

Die in der Einleitung dieses Beitrags angesprochene Hürden bei der Wahl von Lehrveranstaltungen adressiert eine Anwendung, in der Lernende die semantische Ähnlichkeit von Kursen und Studienbriefen mittels Visualisierungen in Form von Chord-Diagrammen, Forced-Layout Graphen und Heatmaps explorieren können. Diese Node-Link Diagramme eignen sich vorrangig für kleine Graphen, da die Visualisierung durch Überschneidungen der Kanten schnell unübersichtlich wird. Die Relationen der vielen Kurseinheiten sind erst dann nachvollziehbar, wenn man sie explorativ aufruft. Die Lesbarkeit der Heatmap hängt maßgeblich von der Anordnung der Elemente ab.

Das zweite Anwendungsszenario richtet sich vordergründig an Lehrende und Autoren von Lernmaterialien. In einem Learning Analytics Dashboard [Sc16] werden Kursbelegungsstatistiken mit den semantischen Relationen verknüpft. Das Dashboard setzt sich aus sechs Kacheln in einem dreispaltigen Layout zusammen: (1) In einer Adjazenzmatrix ist die Ähnlichkeit der im Kurs enthaltenden Studienbriefe abgebildet (Abb. 4, links). (2) Die fünf ähnlichsten Kurse sind ebenfalls in einer Matrix dargestellt (Abb. 4, mitte). (3) Ein Liniendiagramm zeigt den Verlauf der Kursbelegungen der letzten Jahre (Abb. 4, rechts). Darüber hinaus beinhaltet das Dashboard eine Statistik der am häufigsten (4) zuvor, (5) gleichzeitig und (6) im Anschluss belegten Kurse in Form von horizontalen Balkendiagrammen.

Im dritten Anwendungsfall wurde eine adaptive Navigationsunterstützung im Sinne einer *direct guidance* [Br07] entwickelt. Als Komponente einer personalisierten Lernumgebung werden für einen gegebenen Textabschnitt, jene Studienbriefe ermittelt, die die größte semantische Ähnlichkeit zum Text aufweisen. Die Verfahren der *Word Embeddings* und *Document Embeddings* werden hier kombiniert. Als Verweis auf das vorhandene Vorwissen können jene Studienbriefe als Link ausgeben werden, welche die Lernenden bereits belegt haben oder noch nicht belegt wurden. Auf diese Weise lassen sich innerhalb eines Studiengang, einer Fakultät oder einer Hochschule weiterführende Lehrveranstaltungen zu einem bestimmten Thema identifizieren und bewerben. Um jedoch keine Kurse mit einer zu geringen semantischen Ähnlichkeit anzubieten, ist es wichtig einen Schwellwert für den Ähnlichkeitsgrad zu definieren.

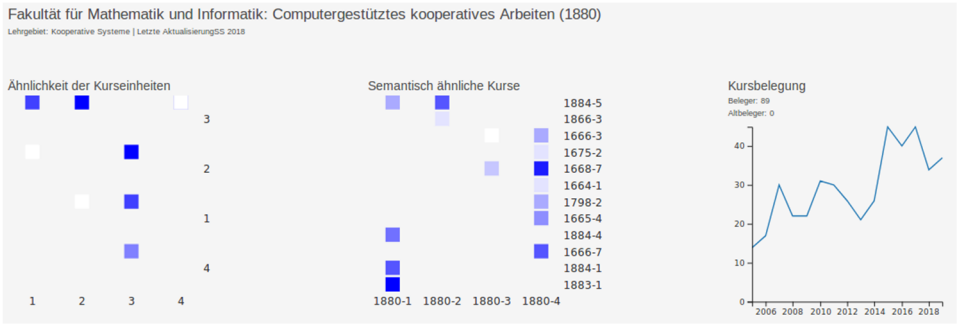



Abb. 4: Ausschnitt aus dem Dashboard für Lehrende

## 5. Fazit und Ausblick

Es wurde ein erweiterbarer Korpus der Fakultät Mathematik und Informatik der FernUniversität in Hagen erstellt. Dabei wurde besonderer Wert darauf gelegt, dass dieser ohne manuellen Aufwand erweiterbar skalierbar ist. Der Korpus ermöglicht einen speichereffizienten Zugriff auf einzelne Studienbriefe oder auf mehrere Studienbriefe je Fakultät, Lehrstuhl und Kurs, sodass er als Grundlage weiterer Studien dienen kann. Anschließend wurden Methoden zur *Feature Extraction* der Dokumente untersucht. Dabei wurde die Abbildung der Semantik in der Vektorrepräsentation fokussiert. Für das ausgewählte PV Modell aus [LM14] wurde gezeigt, dass PV selbst bei Texten mit mehreren tausend Wörtern semantische Informationen abbilden können. Die Ergebnisse wurden mit einem Goldstandard evaluiert und weisen eine hohe Korrelation dazu auf. In Relation zu vergleichbaren Arbeiten (z.B. [Ce17]) wurden im vorliegenden Beitrag wesentlich größere Texte mit mehreren tausend Sätzen, statt nur einzelner Sätze miteinander verglichen, welche präziser semantisch zugeordnet werden können. Durch die *Word* und *Document Embeddings* lässt sich die Ähnlichkeit zweier Kurse gegenüber den Anwendern des Systems durch die Betrachtung der jeweils untergeordneten zu einem Kurs gehörenden Studienbriefen begründen. Durch eine kapitelgenaue Analyse könnten in einem nächsten Entwicklungsschritt die Relationen der Studienbriefe anhand der Relationen der enthaltenen Kapitel nachvollziehbar gemacht werden. Um die Reliabilität der Evaluation zu verbessern, sollte der Goldstandard erweitert werden.

Der hier vorgestellt Ansatz unterliegt einigen Einschränkungen, die gleichzeitig einen weiteren Forschungsbedarf anzeigen. Im Zusammenhang mit *Documents Embeddings* wurden bislang intrinsische Informationen zum Inhalt der Dokumente nicht betrachtet (vgl. [DOL15] sowie LDA oder LSA). Auch Homonyme blieben bislang unberücksichtigt, könnten jedoch durch gelabelte Texte gelernt und auf andere Texte angewendet werden. Um die Lernmaterialien eines Kurses vollständig im Korpus abbilden zu können, sollten auch Texte aus Schaubildern und anderen Visualisierungen einbezogen werden. Die in Abschnitt 4 aufgezeigten Anwendungsmöglichkeiten verdeutlichen mögliche Einsatzfelder für die Nutzung der semantischen Relationen von Studientexten, erfordern jedoch noch weitere Untersuchungen. Insbesondere ist zu prüfen, ob die in der Einleitung genannten Ziele damit erreicht werden können.

In dieser Arbeit wurde exemplarisch gezeigt, wie sich die STS von umfangreichen textuellen Lernressourcen einer Fernuniversität untersuchen lässt. Die Methoden sind jedoch auch auf Präsenzhochschulen übertragbar, welche u.a. mehr mit Präsentationsfolien arbeiten. Darüber hinaus ist es vorstellbar, Kurse und Studiengängen verschiedener Hochschulen miteinander zu vergleichen [Br16] und damit die Studienplatzwahl zu erleichtern. Aus der administrativen Perspektive der Studiengangsplanung und Akkreditierung könnten sich weitere Einsatzfelder der Technologie ergeben. Dies funktioniert nur, insoweit textuelle Repräsentationen von Lernmaterialen wie etwa Foliensätze, Videotranskripte oder Online-Kurse den Inhalt einer Lehrveranstaltung abdecken.

## Literaturverzeichnis


[AC18] Askinadze, A.; Conrad, S.: Development of an Educational Dashboard for the Integration of German State Universities' Data. In: Educational Datamining. S. 527–528, 2018.

[Ag12] Agirre, E. et al.: SemEval-2012 task 6: a pilot on semantic textual similarity. In: Proc. Of the first joint conf. On lexical and computational semantics. S. 385-393, 2012.

[AL13] Aher, S.B.; Lobo, L.M.R.J.: Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data. Knowledge-Based Syst. 51, S. 1–14, 2013.

[BE15] Baumann, A; Endraß, M.A.A.: Visual Analytics in der Studienverlaufsplanung. In: Mensch und Computer 2015 Tagungsband. S. 467–469, 2015.

[BLK09] Bird, S.; Loper, E.; Klein, E.: Natural Language Processing with Python. O'Reilly Media Inc., 2009.

[BOH11] Bostock, M.; Ogievetsky, V.; Heer, J.: D3 Data-Driven Documents. {IEEE} Trans. Vis. Comput. Graph. 17, S. 2301–2309, 2011.

[Br07] Brusilovsky, P.: Adaptive Navigation Support. In: The adaptive web. S. 263–290, 2007.

[Br16] Brackhage, C. et al.: ModuleBase: Hochschulübergreifende Datenbank von Studiengangsmodulen, https://github.com/nise/moduleBase, 2016.

[Ce17] Cer, D.; Diab, M.; Agirre, E.; Lopez-Gazpio, I.; Specia, L.: SemEval-2017 Task 1: Semantic Textual Similarity - Multilingual and Cross-lingual Focused Evaluation, 2017.

[DJ13] D'Aquin, M.; Jay, N.: Interpreting Data Mining Results with Linked Data for Learning Analytics: Motivation, Case Study and Directions. In: Learning Analytics and Knowledge (LAK'13). S. S. 155–164, 2013.

[DOL15] Dai, A.M.; Olah, C.; Le, Q. V: Document Embedding with Paragraph Vectors. CoRR. Abs/1507.0, 2015.

[Ga17] Gábor, K. et al.: Exploring Vector Spaces for Semantic Relations. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. S. 1814–1823, 2017.

[Gr09] Grefenstette, E.: Analysing Document Similarity Measures, https://www.cs.ox.ac.uk/publications/publication3348-abstract.html, 2009.

[Ka13] Kardan, A.A. et al.: Prediction of student course selection in online higher education institutes using neural network. Comput. Educ. 65, S. 1–11, 2013.

[Ki15] Kiros, R. et al.: Skip-Thought Vectors. CoRR. Abs/1506.0, 2015.

[Li18] Liu, B. et al.: Matching Long Text Documents via Graph Convolutional Networks. CoRR. Abs/1802.0, 2018.

[Li18b] Lin, J. et al.: Intelligent Recommendation System for Course Selection in Smart Education. Procedia Comput. Sci. 129, S. 449–453, 2018.

[LM14] Le, Q. V; Mikolov, T.: Distributed Representations of Sentences and Documents. CoRR. Abs/1405.4053, 2014.

[Mi13] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J.: Efficient Estimation of Word Representations in Vector Space. In: ICLR. , Scottsdalte, Arizona, 2013.

[OGD16] Ognjanovic, I.; Gasevic, D.; Dawson, S.: Using institutional data to predict student course selections in higher education. Internet High. Educ. 29, S. 49–62, 2016.

[PSM14] Pennington, J.; Socher, R.; Manning, C.D.: GloVe: Global Vectors for Word Representation. In: Empirical Methods in Natural Language Processing (EMNLP). S. 1532–1543, 2014.

[Ri18] Rieger, M.C.: Semantische Relationen in Studienbriefen der FernUniversität in Hagen. Universitätsbibliothek Hagen. http://nbn-resolving.de/urn:nbn:de:hbz:708-dh9737, 2018.

[Sc16] Schwendimann, B. et al.: Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Trans. Learn. Technol. PP, 2016.

[SS11] Spasojevic, N.; Poncin, G.: Large Scale Page-Based Book Similarity Clustering. In: ICDAR 2011, 2011.

[Wi53] Wittgenstein, L.: Philosophical Investigations. In: New York: The Macmillan Company. S. 272. Blackwell, 1953.

[WIZ15] Weiss, S.M.; Indurkhya, N.; Zhang, T.: Fundamentals of Predictive Text Mining. Springer London, London, 2015.

[Zh15] Zhu, Y. et al.: Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books. arXiv e-prints, 2015.

[ZS10] Zare-ee, A.; Shekarey, A.: The effects of social, familial, and personal factors on students' course selection in Iranian technical schools. Procedia - Soc. Behav. Sci. 9, S. 295–298, 2010.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 45

# Themenübergreifende Diskursklassifikation auf Basis von Word Embeddings und Sequenzfeatures

**Experimente hin zu einer automatischen Diskursanalyse**

Tim Steuer[1] und Christoph Rensing[2]

**Abstract:** Zur Beobachtung von kollaborativen Lernprozessen ist Diskursanalyse ein hilfreiches Werkzeug. Dazu wird der Textkorpus von Annotatoren händisch segmentiert und die Segmente nach ihrer Funktion klassifiziert. Dies ist zeitaufwendig und kostspielig. Automatische Modelle versprechen Zeitersparnis sowie Echtzeitanalysen des Diskurses. Diese könnten direktes Feedback, beispielsweise durch Visualisierungen, an die Lernenden ermöglichen. Automatische Modelle benötigen jedoch manuell annotierte Trainingsdaten. Außerdem sind sie meist vom Diskursvokabular abhängig und generalisieren schlecht über Themengrenzen hinweg. Die dadurch notwendige, häufige Neuerstellung von Trainingskorpora, verringert die Zeitersparnis durch Automatisierung und macht Echtzeit Analyse unmöglich. In dieser Arbeit wird ein Klassifikationsverfahren basierend auf Word Embeddings und Sequenz Features vorgestellt, welches vier Arten von Diskurssegmenten unterscheidet. Das Verfahren erreicht gute Evaluationsergebnisse, mit einer besseren Klassifikationsgüte als Verfahren aus verwandten Arbeiten (Cohens $\kappa > 0.7$). Außerdem generalisiert das Verfahren, auf dem Korpus, ohne weiteres Training von einem Themengebiet auf ein anderes. Dies würde die Notwendigkeit von themenspezifischen Trainingskorpora stark verringern.

**Keywords:** Diskursanalyse, Word Embeddings, Machine Learning

## 1 Motivation

In modernen E-Learning Szenarien ist textueller Diskurs allgegenwärtig. Lernende nutzen beispielsweise Diskussionsforen oder Chats, um zu diskutieren, sich zu organisieren oder um den Kontakt zu ihren Mitstreitern zu halten. Die so entstehenden Unterhaltungen sind mehr als reiner Informationsaustausch, sondern fördern den Lernprozess. Dies ist eine der Grundannahmen von sozio-konstruktivistischen Lerntheorien und von dem darauf aufbauenden Computer unterstütztem kooperativem Lernen (CSCL). Um zu verstehen, wie Teilnehmende in der Diskussion lernen, gibt es im Forschungsfeld CSCL Kodierungshandbücher, mit deren Hilfe man die unterschiedlichen Funktionen der einzelnen Diskussionsbeiträge sichtbar machen kann. Dazu wird jeder einzelne Diskussionsbeitrag zunächst segmentiert und die einzelnen Segmente werden dann jeweils ihrer Funktion

[1] Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstraße 10, 64283 Darmstadt, tim.steuer@kom.tu-darmstadt.de
[2] Technische Universität Darmstadt, Fachgebiet Multimedia Kommunikation, Rundeturmstraße 10, 64283 Darmstadt, christoph.rensing@kom.tu-darmstadt.de

doi:10.18420/delfi2019_234

nach klassifiziert. Hat man beispielsweise einen Diskussionsbeitrag: „Das war ein guter Punkt, jedoch bin ich nicht der Meinung, dass…“, so könnte man diesen Beitrag in zwei Segmente mit verschiedenen Klassen unterteilen: in ein Segment, „Das war ein guter Punkt“, das man als motivierend klassifizieren könnte und in ein zweites diskursspezifisches klassifizierbares Segment „jedoch bin ich nicht der Meinung, dass…“.

Nach der Annotation kann man mit Hilfe der annotierten Klassen Aussagen über die ablaufenden Gruppenprozesse oder über den Effekt von pädagogischen Interventionen auf das Gruppenverhalten treffen [We06]. Ist eine automatische Segmentierung und Klassifikation in Echtzeit möglich, so kann man über Visualisierungen oder Instruktionen auch kontinuierliches Feedback an die Lernenden geben (Learning Analytics).

Einen Diskussionskorpus manuell zu annotieren ist jedoch sehr aufwendig. Jeder einzelne Beitrag der Diskussion muss, um die Objektivität der Annotation zu gewährleisten, von mindestens zwei Ratern gelesen, segmentiert und dann klassifiziert werden. Streitfälle müssen diskutiert und aufgelöst werden, bis beide Rater eine substanzielle Übereinstimmung erreichen.

Eine computergestützte automatische Annotation des Korpus wäre vorteilhaft, um eine schnelle Annotation von großen Korpora zu ermöglichen oder Korpora fortlaufend während des Kurses zu annotieren.

Aus einer technischen Perspektive handelt es sich hierbei um ein Segmentierungs- beziehungsweise Klassifizierungsproblem. Beiden Problemklassen ist gemein, dass Lösungsansätze typischerweise auf überwachtem maschinellem Lernen aufbauen, also einen manuell vorannotierten Korpus nutzen, um die statistischen Regelmäßigkeiten zu lernen, die charakteristisch für eine Klasse oder ein Segment sind. Anschließend lässt sich das Wissen über die so gelernten Regelmäßigkeiten nutzen, um Datensätze mit ähnlicher Struktur zu segmentieren oder klassifizieren. Die statistischen Muster werden dabei bei Texten in der Regel auf Basis des Vokabulars gelernt (z.B. Bag-of-Words Modelle) [Go17]. Daraus folgt, dass die Repräsentation des Vokabulars ausschlaggebend für den Algorithmus ist. Ändert sich das Vokabular, zum Beispiel, weil man anstelle eines Themas aus der Physik ein Thema aus der Biologie diskutiert, so kommen im Diskurs viele Wörter vor, die nicht in dem Trainingskorpus waren und dass verschlechtert die Klassifizierungsergebnisse des Algorithmus [Go17].

Diese Arbeit leistet nun einen Beitrag zur Lösung der oben genannten Probleme, indem wir einen neuen Klassifizierungsansatz für bereits vorsegmentierte, englische Textpassagen vorstellen. Wir arbeiten mit vier verschiedenen Klassen (Tabelle 1): Off-Topic, Organisation, Motivation und Diskursbeitrag. Alle vier Klassen sind pädagogisch relevant. So ist eine Unterscheidung zwischen Off-Topic und Diskursbeiträgen wichtig, um als Tutor schnell zwischen Diskursbeiträgen, die meine Aufmerksamkeit erfordern, und anderen Beiträgen zu unterscheiden. Gleichzeitig gibt es Evidenz dafür, dass sowohl Motivation als auch Selbstorganisation einer Gruppe wichtige Indikatoren für die Effektivität der Gruppe darstellen [Ro18] [Xi11].

| Klasse | Beispielsegment aus Korpus |
|---|---|
| Diskursbeitrag | In my view metacognition, that is, being able to understand, analyze, and control one's cognitive processes, is important in any kind of learning from individual (solo) learning to collaborative and CSCL. Thus, metacognitive knowledge and skills are necessary for effective learning. |
| Off-Topic | "A problem shared is a problem halved, a joy that's shared is a joy made double.", very nice quote indeed. I would like to express another personal quote like saying, "A problem shared is a problem doubled" (p.s. this is of course personal view in the example of some countries in the modern world, some countries share their problems and as a result the problem in some cases is doubled not solved :) ) |
| Organisation | What if we start simply by brainstorming: why is metacognition important or unimportant in CSCL? |
| Motivation | you are awesome! danke :) |

Tab 1: Beispiele für die vier unterschiedenen Textklassen aus dem verwendeten Korpus

Nachfolgend präsentieren wir ein Verfahren, das folgende Beiträge zum bestehenden Stand der Forschung leistet:

- Das Verfahren kann vier pädagogisch wichtige Klassen mit einer für empirische Studien genügenden Güte (Cohens $\kappa > 0.7$) voneinander abgrenzen.
- Das Verfahren verbessert die Klassifizierungsgüte im Vergleich zu in verwandten Arbeiten verwendeten Standardverfahren gemessen an $F_1$ Maß sowie Cohens $\kappa$.
- Das Verfahren benutzt eine vorab trainierte Vektorrepräsentation von Wörtern, sogenannte Word Embeddings, um keine direkten Abhängigkeiten zum Trainingsvokabular aufzubauen. Aufgrund dessen vermuten wir eine bessere Generalisierbarkeit über Diskursthemengrenzen hinweg. Wir präsentieren erste Evaluationsergebnisse, die auf den Erfolg solch einer themenübergreifende Generalisierung hindeuten.

## 2 Verwandte Arbeiten

In den letzten Jahren wurden verschiedene Verfahren zum automatischen Klassifizieren von Forumsdiskussionen präsentiert. Je nach Zielsetzung arbeiten die Verfahren dabei mit verschiedenen Klassen und klassifizieren Segmente, Beiträge oder Threads. Nachfolgend beschreiben wir prototypisch einige Verfahren und unterscheiden im Folgenden zunächst

danach, ob das Verfahren ohne weiteres Training auf andere Korpora generalisiert oder nicht.

Ein Verfahren, das auf ganzen Forenposts arbeitet und ohne Trainingskorpus auskommt, wird von Ezen-Can et al. vorgestellt [Ez15]. Die Autoren arbeiten mit 550 Posts von 155 Lernenden in einem MOOC und nutzen den k-medoids Clustering Algorithmus, um Posts nach Dialogakten zu gruppieren. Anschließend verwenden sie Latent Dirichlet Allocation, um die Cluster mit Stichwörtern zu beschriften. Während das Verfahren aufgrund seines trainingslosen Clustering-Algorithmus leicht zu generalisieren ist, unterscheidet es sich von dem hier vorgestellten Verfahren vor allem in zwei Aspekten. Erstens, die Autoren können nicht a priori festlegen, welche Gemeinsamkeiten die gefundenen Gruppen haben, denn dies hängt vom Clustering auf dem konkreten Korpus ab. Zweitens arbeitet das Verfahren auf ganzen Posts und nicht auf Segmenten.

Ein Korpus unabhängiges Verfahren, das nicht auf Segment- oder Postebene arbeitet, sondern ganze Threads klassifiziert, wurde von Rossi und Gnawali entwickelt [Ro14]. Die Autoren nutzen zwar überwachte Klassifikation, achten bei ihren Eingabedaten aber darauf, dass diese nicht von der Textsprache abhängen. Stattdessen werden quantitative Werte wie beispielsweise die Anzahl der Wörter oder das soziale Netzwerk der im Thread Beteiligten herangezogen. Die Autoren klassifizieren sechs Klassen (General Discussion, Assignments, Meetups, Lectures Logistics und Feedback). Dabei erreichen sie je nach Klasse ROC AUC Werte zwischen 0.58 und 0.89. Das Verfahren unterscheidet sich vom hier präsentierten vor allem darin, dass komplett auf Informationen aus dem Vokabular verzichtet wird. Weiterhin arbeitet es auf ganzen Threads anstelle von Segmenten.

Weiterhin gibt es Verfahren, die vom Vokabular des Trainingskorpus abhängig sind. Wise et al. haben eine automatische Klassifikation von MOOC Threads vorgestellt, die Off-Topic von themenspezifischen Threads abgrenzt [Wi16]. Dabei benutzen die Autoren Bag-of-Words Features mit Unigrammen sowie Bigrammen und eine Support Vector Machine (SVM) als Klassifikationsverfahren. Die Autoren erreichen auf ihren zwei Klassen eine Treffgenauigkeit von 0.86 und ein Cohens $\kappa$ von 0.64 und können somit drei-Viertel aller Threads in die korrekte Klasse einteilen. Die Unterschiede zu diesem Artikel liegen darin, dass die Autoren ihr Verfahren direkt an das Vokabular der Threads koppeln und dass die Autoren nur zwei Klassen unterscheiden.

Die Arbeit von Rosé et al. ist Korpus abhängig und arbeitet auf Segmenten [Ro08]. Die Zielsetzung der Autoren war es herauszufinden, welche Ansätze vielversprechend für die komplette Automatisierung eines Kodierungshandbuches sind. Dabei beschäftigt sich die Arbeit zu einem Großteil mit der Klassifizierung in den sieben Dimensionen des Kodierungshandbuches von Weinberger und Fischer [We06]. Die Autoren benutzen eine Vielzahl von typischen Features wie Unigramme, Bigramme oder Satzzeichen aber konstruieren auch neue, für Segmente wichtige Features wie die Ähnlichkeit zu den Vorgängersegmenten, die Posttiefe oder die Klassen der Vorgängersegmente. Eine Evaluation dieser neuen Features zeigt, dass diese die Klassifizierungsgüte signifikant verbessern. Außerdem testen die Autoren verschiedene Klassifizierungsalgorithmen auf ihren

Daten, unter anderem Algorithmen, die explizit auf die sequenzielle Struktur von Segmenten ausgelegt sind. Die Ergebnisse zeigen jedoch, dass die Wahl des Klassifizierungsalgorithmus nicht ausschlaggebend für die Klassifizierungsgüte ist und es keine signifikanten Unterschiede zwischen einem SVM Ansatz und spezialisierten Algorithmen gibt. Die Arbeit unterscheidet sich von der hier vorgestellten durch ihren Fokus auf ein komplettes Kodierungshandbuch und durch ihre direkte Abhängigkeit vom Vokabular über die Unigramm und Bigramm Features.

## 3 Begründung und Beschreibung unseres Ansatzes

Das konkrete Klassifizierungsproblem befindet sich im folgenden Kontext. Gegeben ist ein mit vier Klassen vorannotierter, englisch sprachiger Korpus auf Segmentebene. Gesucht ist ein Verfahren, das es ermöglicht, einem nicht im Korpus enthaltenem Segment die richtige Klasse zuzuweisen. Dabei soll es möglichst gut generalisieren, also unabhängig von der konkreten Wortwahl des Korpus sein.

Dabei ist der hier verwendete Ansatz nicht unüberwacht, wie beispielsweise das Verfahren von Ezen-Can et al., da die Klassen a priori definiert sind und unbekannte Segmente in derselben Art und Weise annotiert werden sollen, wie bereits bekannte. Außerdem verwendet der hier gezeigte Ansatz sprachabhängige Features, im Gegensatz zu Rossi und Gnawali, da davon auszugehen ist, dass das Vokabular für die Klassen im Korpus wichtige Information beinhaltet. Grundlegend stützt sich unser Klassifizierungsalgorithmus auf den Einsatz von vortrainierten Word Embeddings zusammen mit einer SVM (RBF Kernel). Sowohl Rose et al. als auch Wise et al. haben gezeigt das eine SVM auf ähnlichen Klassifizierungsproblemen gute Ergebnisse mit Bag-of-Words Features erzielen kann. Es sind allerdings verschiedene Schwächen der Bag-of-Words Ansätze bekannt:

Die meisten Klassifikationsalgorithmen, so wie auch SVMs, erreichen bessere Ergebnisse, wenn die Anzahl der Features deutlich kleiner als die Anzahl der Trainingsdaten ist [Ha09]. Bei Bag-of-Words-Ansätzen wird jedes Wort im Vokabular als One-Hot kodierter Vektor φ dargestellt. Der Bag-of-Words (BOW) eines Textsegmentes *S* ist dann eine Linearkombination dieser Wortvektoren wobei man $a_{\mathrm{i}} = 1$ im einfachsten Falle wählt:

$$BOW_{segment} = \sum_{\varphi_i \epsilon S} a_i \varphi_i \tag{1}$$

Dabei wächst die Dimension des Feature-Vektorraumes linear mit der Anzahl an unterschiedlichen Wörtern im Vokabular. Dies führt bei größeren Texten oftmals zu einer sehr großen Anzahl Features und dies beeinflusst die Güte der Klassifikation nachteilig. Deshalb versucht man in diesen Ansätzen oftmals das Vokabular zu beschränken. Dazu filtert man zum Beispiel alle Wörter, die sehr selten oder sehr häufig in einem Text vorkommen heraus, um nur noch für das Trainingsset aussagekräftige Wörter zu behalten [Go17]. Dies verschlechtert allerdings dann wieder die Generalisierbarkeit des Klassifikationsalgorithmus.

Ein weiterer Nachteil von Bag-of-Words ist, dass Wörter, die eigentlich semantisch ähnlich sind, in ihrer Vektorrepräsentation orthogonal zueinanderstehen. Hat man beispielsweise die Wörter "jedoch" und "wohingegen" besitzen die Vektoren den gleichen Abstand wie „jedoch" und „Hund". Der Klassifikationsalgorithmus kann deshalb die semantische Ähnlichkeit der Eingaben nicht ausnutzen. Verwendet man Word Embeddings anstelle von Bag-of-Words werden diese Nachteile abgemildert:

Die Technik der Word Embeddings wurde ursprünglich im Zusammenhang mit neuronalen Netzen entwickelt und erlaubt, wie auch die Bag-Of-Words Ansätze, Wörter als Vektoren zu repräsentieren und als Features in Klassifizierungsalgorithmen zu verwenden [Co11]. Die Konstruktionsvorschrift der Embeddings betrachtet jedoch nicht jedes Wort als eine separate Dimension, sondern das Wort in seinem Kontext [Co11]. Bei der Konstruktion wird eine fixe Dimension für die Embeddingvektoren festgelegt und die einzelnen Wörter werden dann in diese Dimension eingebettet [Co11]. Wählt man zum Beispiel eine Dimension von 300, wird für jedes Eingabewort ein Vektor der Dimension 300 berechnet. Durch die Art, wie während der Einbettung der Kontext einbezogen wird, liegen Vektoren von Wörtern, die semantisch ähnlich sind, dicht beieinander.

Um nun einen Text *S* mit *k* Wörtern mithilfe dieser Embeddingvektoren zu repräsentieren, berechnen wir einen Continuous-Bag-of-Words (CBOW) [Go17] erneut als Linearkombination der Embeddingvektoren, wobei *f* eine Abbildung ist, die jedem Wort *w* des vortrainierten Vokabulars seinen n-dimensionalen Embeddingvektor zuordnet, im einfachsten Fall mit $a_i = 1$:

$$\mathrm{CBOW}_{\mathrm{segment}} = \frac{1}{k} \sum_{w \in S} a_i f(w) \tag{2}$$

Der Continuous-Bag-of-Words adressiert nun die zwei oben genannten Probleme des Bag-of-Words. Zunächst wächst die Dimension des Continous-Bag-of-Words nicht mehr linear mit der Anzahl der unterschiedlichen Wörter im Vokabular, sondern ist konstant, was vorteilhaft für den Klassifikationsalgorithmus ist. Außerdem besitzen ähnliche Wörter, aufgrund der Konstruktion, ähnliche Vektoren. Dadurch liegen Texte aus semantisch ähnlichen Wörtern in ihrer Vektorrepräsentation näher beieinander, was die Klassifikation erleichtert. So haben Collobert et al. gezeigt, dass Word Embeddings aufgrund dieser Eigenschaft eine Vielzahl von Sprachverarbeitungstasks verbessern [Co11].

Schließlich ist die Nutzung der Embeddings noch mit einem weiteren wichtigen Vorteil verbunden. Es wurde gezeigt, dass wenn die Embeddings einmal auf einem ausreichend großen Korpus trainiert wurden, ihre semantische Ähnlichkeit auf andere Korpora und somit andere Klassifikationsaufgaben übertragbar ist [Co11]. Durch die Nutzung vortrainierter Embeddings (GloVe Common Crawl 300 [Pe14]), die auf einem sehr großen Kor-

pus trainiert wurden, können deshalb auch Wörter, die nicht im Trainingskorpus vorkommen, später als aussagekräftiges Feature genutzt werden, denn es existieren dafür ebenfalls Embeddingvektoren.

Neben den Embeddings nutzt der hier beschriebene Ansatz einige weitere Features, um den Kontext, in dem die einzelnen Segmente geäußert wurden, zu modellieren. Für jedes Segment werden zunächst einige einfache Kennzahlen betrachtet: Die Posttiefe, die Anzahl der Fragezeichen sowie die Anzahl der direkten Zitate innerhalb des Segmentes. Außerdem werden temporale Äußerungen mithilfe des Temporal Tagger Heideltime [St12] annotiert und gezählt, um die Anzahl der explizit genannten Zeitangaben herauszufinden, zur besseren Klassifikation von Organisationsbeiträgen.

Weiterhin kommen, ähnlich zu Rose et al. Sequenzfeatures zum Einsatz. Dazu werden die Kosinus-Ähnlichkeiten des Continuous-Bag-of-Words des aktuellen Segmentes mit denen der vorherigen verglichen, um eine Änderung des Diskursverlaufs zu detektieren. Außerdem wird das Sentiment des aktuellen Segments, sowie der vorherigen, berechnet, um motivationale Abschnitte besser zu detektieren. Abschließend werden ebenfalls die Klassen der vorherigen Segmente miteinbezogen, da es gut möglich ist, dass es typische Abfolgen von Klassen gibt (z.B. Motivation immer vor Diskurs). Dazu wird Stanford CoreNLP verwendet [Ma14]. Insgesamt erhalten wir so die in Tabelle 2 gezeigten Features. Die genutzte SVM wird mit dem Pythonframework scikit-learn und einem RBF Kernel trainiert. Dabei skalieren wir $\gamma$ automatisch und gewichten die Kostenfunktion mit der Klassenhäufigkeit.

| | **Wortrepräsentation** | **Postkennzahlen** | **Sequenzrepräsentation (jeweils letzte 15)** |
|---|---|---|---|
| Art | GloVe-Common-Crawl Embeddings | Posttiefe, Frageanzahl, Zitatanzahl, Zeitangaben, aktuelles Sentiment | vorherige Sentimente, Klassen, Kosinus Ähnlichkeit |
| Dimension | 300 | 5 | 45 |

Tab. 2: Klassifikationsmerkmale mir ihrer Dimension. Aufteilung in sprachliche Features, segmentspezifische Features und sequenzspezifische Features

# 4 Korpus

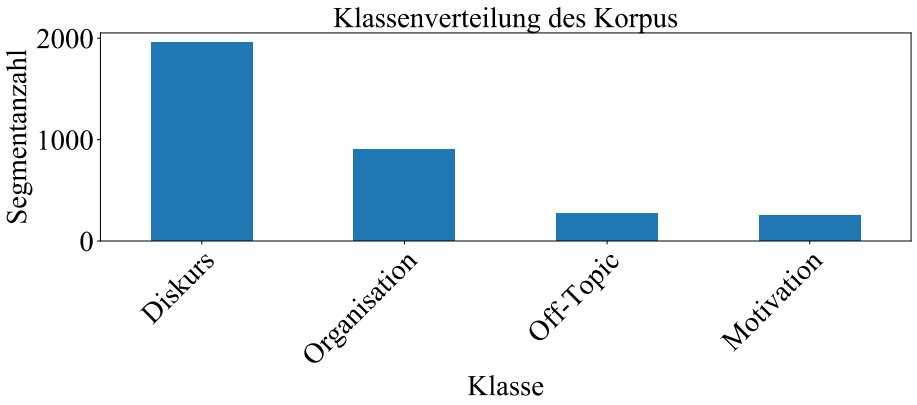



Abb. 1: Verteilung der einzelnen Klassen im Evaluationskorpus

Um unser Klassifikationsverfahren zu evaluieren, nutzen wir einen vorannotierten Korpus. Dieser Korpus wurde während zwei Durchläufen eines sechswöchigen online Universitätskurses dreier Universitäten in den Jahren 2015 und 2016 erstellt. Die Arbeitssprache des Kurses war Englisch und die Studierenden diskutieren jeweils in Kleingruppen aus drei bis fünf Mitgliedern online auf Facebook. Dabei wurde eines von drei Themen jeweils für vier Wochen von den Teilnehmern diskutiert: „CSCL Skripte“, „Motivation und Emotion im Lernen“ sowie „Metakognition“. Die quantitative und qualitative Teilnahme an der Diskussion wurde benotet.

Insgesamt besteht der Korpus aus 3246 einzelnen Segmenten, die von zwei Annotatoren nach einem Kodierungshandbuch annotiert wurden. Die Interrater-Reliabilität betrug dabei Cohens $\kappa = 0.85$. Für die Evaluation wurden die Annotationen auf die vier Klassen Off-Topic, Organisation, Motivation und Diskurs zusammengefasst. Dabei zeigt sich (Abbildung 1) eine starke Ungleichverteilung der Klassen, wobei Beiträge zum inhaltlichen Diskurs dominieren. Nachfolgend wird deshalb als Evaluationsmetrik neben Cohens $\kappa$ ebenso das Makro F1 Maß sowie Precision und Recall angegeben, um die Ungleichverteilung der Klassen zu berücksichtigen.

# 5 Evaluation

Zu evaluieren sind drei Fragestellungen. Erstens: Ist es mit unserem Verfahren möglich, die vier gegebenen Klassen ausreichend gut zu klassifizieren? Wir gehen in Übereinstimmung mit Rose et al. davon aus, dass wir dieses Ziel erreichen, wenn wir ein Cohens $\kappa >$

0.7 erreichen, welches auch in empirischen Studien als beachtliche Übereinstimmung angesehen wird.

Zweitens: Schneidet unser Verfahren besser auf den Daten ab, als die Standardverfahren? Wir testen dazu ob normale Bag-of-Words Verfahren auf den Daten ein niedrigeres F1-Maß erreichen, als das Continuous-Bag-of-Words Verfahren. Dabei vergleichen wir mit einem Bag-of-Words der das gesamte Vokabular benutzt und einem Bag-of-Words, der das TF-IDF Maß nutzt [Go17], um hoch- und niederfrequente Wörter herauszufiltern. Dies ist ein typischer Standardansatz, um das oben skizzierte Problem der hohen Dimensionalität des Bag-of-Words zu lösen.

Drittens: Generalisiert unser Verfahren auf unbekannten Themen besser als Bag-of-Words Ansätze? Wir zerteilen unseren Korpus bei der Evaluation immer entlang der Themengrenzen des Kurses und trainieren mit zwei der drei Themen. Dabei ist die Anzahl der Segmente pro Thema ungefähr gleich groß. Das dritte Thema wird dann als Testmenge genutzt. Sofern unser Verfahren gut generalisiert, sollten auf allen Testmengen die gleichen F1 Maße herauskommen, denn es sollte egal sein, ob der Algorithmus mit Themen (1,2) trainiert wird und auf 3 generalisiert oder mit Themen (2,3) trainiert wird und auf 1 generalisiert. Gleichzeitig sollten die Bag-of-Words Ansätze eine größere Streuung der F1 Maße aufweisen, da sich das Vokabular in den einzelnen Themen unterscheidet und so die Generalisierbarkeit erschwert.

Abbildung 2 zeigt die Evaluationsergebnisse, wobei der Korpus entlang der Themengrenzen geteilt wurde, auf zwei Themen trainiert und auf dem verbleibendem evaluiert wurde. Da Off-Topic Beiträge ungleichmäßig viel in der ersten Hälfte des Kurses vorkommen, haben wir diese gleichmäßig auf die Trainings- und Testmenge verteilt, um zu gewährleisten, dass die einzelnen Klassen in jeder Menge ungefähr gleichermaßen verteilt sind. Bei den Ergebnissen wurde jeweils die Wortrepräsentationsfeatures getauscht, die Postkennzahlen sowie die Sequenzkennzahlen blieben gleich.

Der ungefilterte und ungewichtete Bag-of-Words Modell erreicht ein Makro F1-Maß zwischen 0.59 und 0.65 (Precision: 0.68-0.71 Recall: 0.62-0.66) je nach Datenaufteilung. Dabei nutzt er das gesamte Vokabular der Trainingsdaten, im Schnitt um die 6000 Worte bei jeweils ca. 2000 Trainingsdaten. Das TF-IDF Modell mit Häufigkeitsfilter (0.05 bis 0.95) erreicht Makro F1-Maße zwischen 0.67 und 0.72 (Precision: 0.72-0.74 Recall: 0.64-0.69) wobei nach dem Filter 120 Features verbleiben. Das Continuous-Bag-of-Words Modell erreicht ein konstantes Makro F1-Maß von 0.75 (Precision: 0.78-0.81 Recall: 0.72-0.77) bei einer Embedding-Dimension von 300.

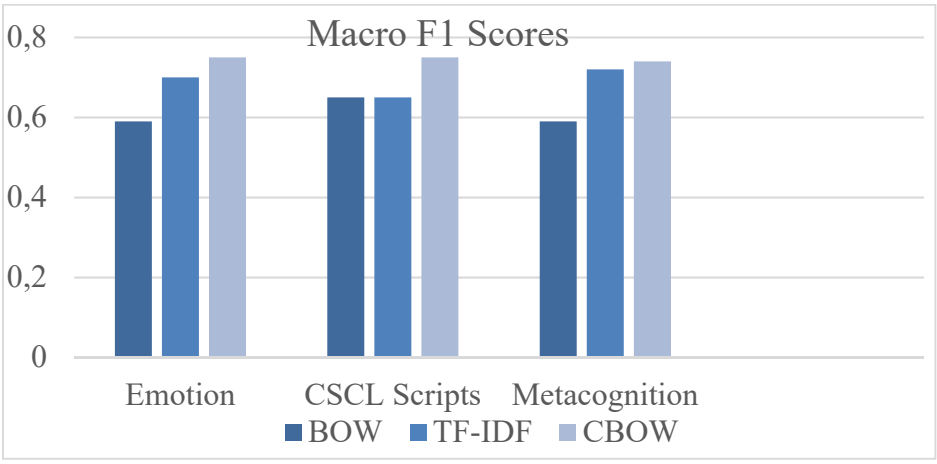



Abb. 2: Evaluationsergebnisse: die Beschriftung einer Balkengruppe bezeichnet die genutzte Testmenge. Die Bag-of-Words SVM beinhaltet das gesamte Vokabular (ca. 6000 Wörter), die TF-IDF SVM nutzt einen Wort Häufigkeitsfilter von min. 0.05 und max. 0.95 (ca. 120 Wörter) und der Continuous -Bag-of-Words SVM nutzt 300-dimensionale GloVe Embeddings.

## 6 Diskussion und Ausblick

Bezüglich der zu evaluierenden Fragestellungen lässt sich folgendes feststellen: Die erste Fragestellung, ob das Verfahren es schafft, die vier Klassen mit ausreichender Güte auseinanderzuhalten, können wir positiv beantworten. Wir haben mit $F1 > 0.7$ und Cohens $\kappa > 0.7$ eine Klassifizierungsgüte erreicht, die den sozialwissenschaftlichen Standards gerecht wird. Zunächst heißt dies, dass es dieses Verfahren einen guten Ausgangspunkt bildet, um Segmentierungsalgorithmen für eine vollautomatische Annotation zu entwickeln, da die prinzipielle automatische Unterscheidbarkeit der Klassen gegeben ist. Im Idealfall bedeutet dies auch, dass man den Klassifizierungsalgorithmus als zweiten Rater eines Korpus betrachten kann und somit nur noch mit einem Rater manuell, statt mit zweien annotieren muss. Diese Idee sollte jedoch in weiteren Arbeiten überprüft werden, denn im Gegensatz zu menschlichen Ratern bringt ein voll automatisierter Algorithmus kein wirkliches semantisches Verständnis über die Daten mit, sondern arbeitet auf den gelernten Mustern, was eine starke Vereinfachung der Semantik darstellt. Dadurch kann sowohl die Reliabilität als auch die Validität des Kodierungsschemas beeinträchtigt werden, was zumindest auf einem Teil der automatisch annotierten Daten wiederum zu testen ist [Ro08].

Die zweite Fragestellung, ob sich ein auf Embeddings basierendes Verfahren auf dem gegebenen Korpus besser schlägt als typische Referenzverfahren aus der Literatur, lässt sich ebenfalls positiv beantworten. Dabei schneidet das auf Embeddings basierende Verfahren in allen Evaluationsszenarien besser ab. Außerdem erreicht nur das Embedding basierte

Verfahren eine sozialwissenschaftlich ausreichende Klassifikationsgüte in allen Aufteilungen. Dabei schneiden die Embeddings auch besser ab als die TF-IDF gewichteten und gefilterten Bag-of-Words Vektoren, obwohl die Dimension des Eingaberaumes, dort geringer war. Wir vermuten, dass dies damit zusammenhängt, dass die TF-IDF Repräsentation hier für die Klassifikation nützliche Wörter verwirft, da diese zu selten oder häufig vorkommen. Um solche Fehler zu vermeiden, kann man komplexere Feature-Engineering Methoden anwenden, und beispielsweise nur die Wörter als Feature verwenden, die auch eine gute Korrelation mit den Ausgangsklassen haben. Dabei besteht immer der Nachteil, dass solche Verfahren die Features abhängig vom Korpus auswählen, was die Generalisierbarkeit verschlechtert. Ein auf vortrainierten Embeddings basierender Ansatz benötigt hingegen überhaupt kein Feature-Engineering und bildet ein großes Vokabular ab. Darum zeigen unsere Ergebnisse, dass vortrainierte Embeddings auch in einer spezialisierten Domäne wie E-Learning eine nützliche Feature Darstellung sein können, die ohne großen Aufwand gute Ergebnisse liefert.

Bezüglich der dritten Fragestellung gibt es Evidenz, dass unser Embedding basiertes Verfahren besser generalisiert als die Bag-of-Words Verfahren. Dabei schwankt das F1 Maß hier bei den verschiedenen Trainings und Testmengen kaum, wohingegen sowohl bei TF-IDF als auch beim klassischen Bag-of-Words mindestens einer der drei Evaluationsdatenpunkte in der Klassifikationsgüte abweicht. Diese Ergebnisse deuten auf eine bessere Generalisierbarkeit durch die Continuous-Bag-of-Words Features hin. Für endgültige Aussagen liegen jedoch zu wenige unterschiedliche Themen im Evaluationskorpus vor. Außerdem unterscheiden sich Trainings- und Testmenge zwar bezüglich ihrer Themen, jedoch kommen die Themen immer noch aus den gleichen Wissenschaftsgebieten. Die Diskussionskultur unterscheidet sich jedoch beachtlich in verschiedenen Wissenschaftsfeldern und es ist nicht klar, ob unsere Ergebnisse, auch auf weiter auseinanderliegende Themen, wie beispielsweise Diskussion über biologische Themen, generalisieren. Weitergehende Arbeiten sollten deshalb mit einem zweiten annotierten Korpus arbeiten, der mindestens ein Thema aus einer anderen Wissenschaftsrichtung behandelt. So könnte man testen, ob die vier Klassen wirklich komplett unabhängig vom diskutierten Themengebiet sind.

Ein wichtiger Anknüpfpunkt für weitere Arbeiten ist die Segmentierung. Zwar kann das hier präsentierte Verfahren die Klassifikation vornehmen, doch benötigt es dazu fertige Segmente. Da ein menschlicher Rater in der Regel beim Segmentieren gleichzeitig klassifiziert, spart das Verfahren im Falle eines einzigen Raters keinen Aufwand. Deshalb ist die Segmentierung der nächste Schritt hin zu einer vollautomatischen Lösung. Außerdem wichtig wäre eine feinere Klassifizierung der einzelnen Segmente wie sie auch in den Kodierungshandbüchern vorkommen, um eine genauere Analyse des Diskurses zu ermöglichen. Dazu muss zunächst evaluiert werden, inwieweit sich die feineren Klassen überhaupt automatisch unterscheiden lassen.

**Danksagung** Wir möchten uns beim Lehrstuhl für Bildungstechnologie und Wissensmanagement der Universität des Saarlandes sowie Eirini Papanastasiu für die Zusammenstellung und Zurverfügungstellung des Korpus bedanken.

## Literaturverzeichnis

[Co11] Collobert, R.; Weston, J., Leon; B., Karlen; M., Kavukcuoglu, K.; & Kuksa, P.: Natural Language Processing (Almost) from Scratch. The Journal of Machine Learning Research, 12, S. 2493-2537, 2011.

[Ez15] Ezen-Can, A.; Boyer, K. E.; Kellogg, S.; Booth, S.: Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. In Proc. of LAK. S. 146–150, 2015.

[Go17] Goldberg, Y.: Neural Network Methods for Natural Language Processing. 2017.

[Ha09] Hastie, T.; Tibshirani, R.; Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2. Auflage, Springer Series in Statistics, 2008.

[Ma14] Manning, C. D.; Surdeanu, M.; Bauer, J.; Finkel, J.; Bethard, S. J.; McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In Proc. of Annual Meeting of the Association for Computer Linguistics: System Demonstrations, S. 55–60, 2014.

[Pe14] Pennington, J.; Socher, R.; Manning, C. D.: GloVe: Global Vectors for Word Representation. In Proc. Of EMNLP. S. 1532–1543, 2014.

[Ro08] Rosé, C.; Wang, Y. C.; Cui, Y.; Arguello, J.; Stegmann, K.; Weinberger, A.; Fischer, F.: Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. International Journal of Computer-Supported Collaborative Learning 3/3, S. 237–271, 2008.

[Ro14] Rossi, L. A.; Gnawali, O.: Language independent analysis and classification of discussion threads in Coursera MOOC forums. In Proc. of IEEE IRI. S. 654–661, 2014.

[Ro18] Romero, M.; Lambropoulos, N.: Internal and External Regulation to Support Knowledge Construction and Convergence in Computer Supported Collaborative Learning (CSCL). Electronic Journal of Research in Education Psychology 9/23, S. 309–330, 2018.

[St12] Strötgen, J.; Gertz, M.: Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards. In Proc. of LREC, S. 3746–3753, 2012.

[We06] Weinberger, A.; Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers and Education, 46/1, S. 71–95, 2006.

[Wi16] Wise, A. F.; Cui, Y.; Vytasek, J.: Bringing order to chaos in MOOC discussion forums with content-related thread identification. In Proc. of LAK. S. 188–197, 2016.

[Xi11] Xie, K.; Ke, F.: The role of students' motivation in peer-moderated asynchronous online discussions. British Journal of Educational Technology 42/06, S. 916–930, 2011.

# Stakeholder attitudes towards digitalization in higher education institutions

Daniel Biedermann[1], Lea Kalbfell[2], Jan Schneider[3] and Hendrik Drachsler[4]

**Abstract:** We report the results of a group concept mapping study on the topic of digital transformation in German higher education. Students, teachers and administrative employees participated in the study where they first gathered statements which were then clustered and ranked according to importance and feasibility. The results show agreement between all groups that matters of culture and mentality remain the biggest factors for a successful digitalization of higher education. While we found agreement between groups on the importance of topics, those working at the institution are more pessimistic of cultural changes than students. The unification of service platforms and the development of validated didactical concepts for digital learning offerings are also seen as open challenges.

**Keywords:** Digital Transformation, Group Concept Mapping, Digitalization

## 1 Introduction

Major changes in education are often slow and face resistance. The stakes are high, stakeholders vary in their backgrounds and may have different and competing interests. There are usually few, if any, absolute truths. An area that is currently transforming education on a large scale is the digitalization of educational institutions. Digitalization touches several sensitive areas, such as automatization and privacy. This sensitivity is especially high in Germany [TM07]. The population is historically very wary of issues regarding privacy, and the GDPR, as well as several privacy scandals have recently made the public even more aware of this topic [TM07].
Thus, to advance digitalization of educational institutions, it is important to be able to advance sensibly, and to know about the attitudes and expectations towards this topic from all the stakeholders that are involved in the process. Stakeholders are the students, the teachers, the administration and - in higher education - the researchers. The attitudes and expectations of all those groups must be considered to successfully commence changes on a large scale. Without knowing what the actual problems are, those that try to push digitalization may not receive all the support that they require. Groups that feel left behind because their issues are not considered could resist change and maybe rightfully so. It is therefore of central importance to obtain as comprehensive an image as possible of the attitudes towards digitalization and e-learning at educational institutions.

[1] DIPF | Leibniz Institute for Research and Information in Education, biedermann@dipf.de
[2] Lea.kalbfell@gmail.com
[3] DIPF | Leibniz Institute for Research and Information in Education, schneider.jan@dipf.de
[4] DIPF | Leibniz Institute for Research and Information in Education, drachsler@dipf.de

doi:10.18420/delfi2019_332

Previous works have used primarily surveys to obtain the attitudes towards specific topics. Ifenthaler et al. have asked students at the University of Mannheim about their attitudes towards Learning Analytics (LA) [IS16]. They have found that students expect personalized systems but are at the same time rather conservative towards data sharing. This slightly contradictory expectation shows that it is important to know the exact conditions that are acceptable for learners. They complemented their survey with a qualitative study on 20 students where they validated their results [SI18]. Schmid et al. performed a large survey in 34 German universities and asked 2759 students, 542 teachers and 84 administrative employees about the state of digitalization in their higher education institution [Sc17]. They asked questions about how widespread digital learning technologies are implemented and about the stakeholders' attitudes towards those technologies. They found that while there appears to be already solid well-rated infrastructure and technology in place, that "...the educational potential of digitalization often remains untapped". In terms of strategic considerations, they identified two major camps that they call "digital advocates", as well as "analog skeptics", with quite opposing views on the furthering of e-learning. Major challenges that remain are legal issues as well as cost [Sc17]. For the student body, they note that students are not necessarily overly competent nor particularly well-versed in utilization of e-learning resources. They seem to primarily use the resources which are made available by the teaching personnel, which in turn seems very heterogeneous in their use of digital learning materials [Sc17].

While surveys offer a good overview, they may not be enough to adequately reflect the differentiated opinions of those involved. There are ideas and desires that may fly under the radar and thus if it is not asked, there will be no response to it. A way to get a perspective that is generated by participants themselves, while still quantifiable is the mixed methods Group Concept Mapping (GCM) methodology. The GCM methodology has previously been used in other studies to get opinions about digital transformations in other regions. A study by Wopereis et al. has investigated failure and success factors for the use of information and communication technologies in the Netherlands [Wo05]. They have found that "people-issues", i.e. participation, information, communication and stakeholder involvement are the most important factors. In another Dutch study, Drachsler et al. have performed a GCM study with 39 participants to get perspectives on Learning Analytics (LA) [Dr14]. While they also captured statements regarding stakeholder attitude, they also found several areas that are more concerned with concrete technologies that should be utilized. Similar was found in an Australian study [Co16], which observed a both attitudes and specific technology instances as relevant.

For our study, we performed a GCM at a German higher education institution to get attitudes and expectations towards digitalization in education. We recruited students, teachers, administrative staff, as well as researchers to obtain their statements regarding the success factors for digital transformation. Participants then grouped and ranked the statements according to importance and feasibility. In the following sections, we present

the GCM methodology and describe how we implemented it. We then reports on the results that we obtained in terms of their descriptive statistics. Finally, we discuss the results and offer our interpretation of them.

# 2 Method

## 2.1 Procedure

Group Concept Mapping (GCM) is a mixed method approach to capture results of collaborative decision-making processes. The goal is for a group to arrive at shared visions about a topic (in this case, the topic of digitalization in education). In contrast to other methods of collecting and analyzing opinions, in GCM the participants themselves generate the structure and ideas around the topic. Using multivariate statistics, the ideas of the participants are analyzed and clustered to objectively identify patterns in the data. The results of this analysis are visualized to allow further interpretation. Details and example of how GCM works and can be applied can be found at e.g. [Dr14] [RK12].

The participants were invited to take part in the study using a dedicated online GCM platform. Participants who finished all phases of the study were rewarded with 50€. Participants were asked about their position at the institution and their prior experience with learning technologies.
For the first phase, the brainstorming phase, participants were asked to complete the statement

*"One success factor for digital education at the [institution] is..."*

with their own ideas. This phase was active for ten days, so participants had several days to generate ideas and they could return at any time during that period to generate further ideas. Statements that were generated were cleaned up by removing duplicates and correcting spelling mistakes.
In the second phase, participants were asked to group the ideas from the first phase based on similarity and assign them a common label. For this task, all the statements were given in one large list and they could drag and drop them into containers that they created. They could create as many containers as they wanted and were requested to avoid categories such as "misc." or "other".
Finally, in the third phase, participants were asked to rate all the items on a 1-to-5 scale for their importance (1 = not important at all; 5 = very important) and feasibility (1 = not feasible at all; 5 = very feasible).

For the analysis, the ideas are represented as points on a two-dimensional plane. Points are closer on this plane if they have been put in the same group more often. The points were then clustered using hierarchical cluster analysis. Cluster suggestions are created beginning with a high number of clusters first, then reducing the number of clusters. Criterion for the quality of the clusters is the average bridging value (BR) of each

cluster. A low bridging value for an item indicates that it has been grouped with other items around it more often. The cluster coherence was furthermore inspected visually. The final number of clusters was decided on in a separate session where multiple researchers agreed on a final number of clusters based on their numerical and visual coherence.

### 2.2 Participants

Participants could complete only some of the steps if they chose to do so. Thus, the number of participants varied between the phases. For the brainstorming phase, 101 participants contributed. The sorting phase was started by 55 and was completed by 46 participants. The importance rating was started by 50 participants and was completed by 44 participants. The feasibility rating was started by 49 participants and completed by 43 participants. Table 1 shows the distribution regarding the position at the institution (multiple mentions were possible) and Table 2 the prior experience with digital learning technologies. The surveys were presented before the sorting phase.

| Position | n | % |
|---|---|---|
| Student | 23 | 28.40 |
| Employed at Chair | 16 | 19.75 |
| Researcher | 15 | 18.52 |
| Teacher | 15 | 18.52 |
| Administrative Staff | 10 | 12.35 |
| Management | 1 | 1.23 |
| Other | 1 | 1.23 |

Tab. 1: Participants position at the institution. Multiple mentions possible

| Experience | n | % |
|---|---|---|
| No prior experience | 2 | 4.44 |
| 1-3 years | 18 | 40.00 |
| 3-6 years | 7 | 15.56 |
| 6-9 years | 9 | 20.00 |
| 10-13 years | 3 | 6.67 |
| More than 13 years | 6 | 13.33 |

Tab. 2: Participants prior experience with learning technologies

## 3 Results

### 3.1 Clusters

We arrived at ten clusters. Their graphical representation can be seen in figure 1. Descriptive statistics of the clusters can be found in table 3. In this section, we report on the clusters that were created and the statements that these clusters contain. For each cluster, we report on those items which can be considered “low hanging fruit”, i.e. those items which have both a high average feasibility (FE) and as a high average importance (IM) rating. We furthermore report on statements that stick out. Wherever possible, we attempt to include the actual quotation in the text

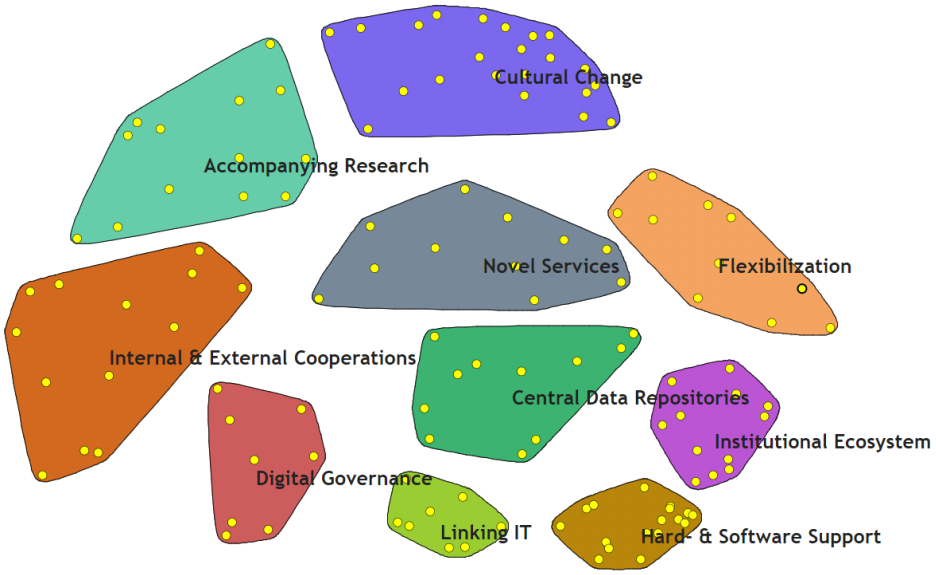



Fig. 1: The final clusters. Yellow dots are the statements within the clusters

Statements in the *Cross-linking of the IT infrastructure* cluster deal with issues and requests towards the various IT services at the institution. Requests include development of central platforms for needs such as jobs and internships (IM = 3.11; FE = 3.91), as well as having the institutional services linked to external platforms and services, such as calendars or certificates (IM = 2.95; FE = 3.14). Furthermore, "...having a central contact person for technical and didactical questions" (IM = 3.70; FE=3.60) is also requested. Ranked as important, yet perceived as not very feasible, are "...permanent positions in ALL areas of work that are responsible for educational innovations - networking and consistent change management" (IM = 3.64; FE=2.58).

*Central Data Repositories* contains statements for which a solution requires means of generating and storing data and making it available for the stakeholders. These include "...the comprehensive provision of literature online (e.g. e-books) also for use outside the university" (IM = 3.63; FE = 3.30) and "... that the competent support regarding legal factors (e.g. scans of literature etc.) is available." Important, yet seen as comparatively very unfeasible are statements regarding utilization of existing data for the purpose of data analytics (IM = 3.61; FE = 2.47).

The cluster *Novel Services* contains expectations and suggestions how e-learning services and platforms could improve education. While the average ratings are particularly unfavorable towards "...Establishing novel degrees, e.g. Nanodegrees, Micromaster (4-week course for a topic)" (IM = 2.50; FE = 2.19), increases in flexibility

through recordings of lectures and seminars (IM = 3.42; FE = 3.53), self-regulated structuring of the studies (IM = 3.79; 3.63,) and the option to reduce presence in lectures by means of e-learning (IM = 3.49; FE = 3.74) are received favorable. The highest ratings are for the statement "...development of didactic concepts for digital education" (IM = 4.23; FE = 3.98).

*Digital Governance* (BR = 0.68) contains requirements for the organization, such as creation of rulesets and the cooperation between divisions within the educational institution. The most important statement in this cluster is "...Improved cooperation of the central institutions for digital education" (IM = 4.16; FE = 3.37) and in a similar vein "...better networking of all employees involved in the IT infrastructure" (IM = 4.00; FE = 3.49), as well as "...that existing information is made transparent and easily accessible" (IM = 3.77; FE = 3.79).

In the cluster *Internal and External Cooperation* are statements which deal with topics of communication and cooperation within the institution. While this cluster has the highest bridging value of all the clusters, indicating that the items have often not been placed in the same category, inspection of the actual items reveals that they in fact mostly share this topic. The common sentiment is that different departments should communicate regularly and work towards building competencies. Training and education are mentioned as ways to achieve this, e.g. "...regular training and further education for teaching staff in the use of available media." (IM = 4.09; FE = 3.81).

In the *Critical Accompanying Research* cluster are those responses which note a need for (continuous) evaluation of the change process. Responses focus on didactics and e.g. demand that "...Lecturers (TAs and Professors) receive didactic training on how to correctly and effectively apply digital education" (IM = 4.16; FE = 3.93) and "digital education is accompanied by research questions/research projects from the areas of media didactics and educational psychology" (IM = 3.80; FE = 3.88).

*Cultural Change* is the cluster which contains the most items (n = 22). Items in this cluster mention attitudes and mindsets that participants see in need of (re-)evaluation for digitalization to become successful. Some statements describe worries, e.g. "...understanding eLearning as a complement for teachers instead of as a substitution" (IM = 3.95; FE = 3.79) and "...despite all enthusiasm for new technologies, do not forget to keep the people the center of attention of strategic considerations" (IM = 3.89; FE = 3.95). Meanwhile some statements take the opposite stance and demand openness "to admit ideas first, not to nip things in the bud at the very beginning because of data protection or finances" (IM = 3.84; FE = 3.65). Lastly, statements in this cluster also deal with the transition and fusion of digital and traditional means of teaching. The most important and most feasible statement is "…if digital and analogue education are combined and interlinked in a meaningful (didactically valuable) way (keyword: blended

learning)" (IM = 4.23; FE = 4.00). This cluster is also the one with the highest average importance (3.78).

In the *Flexibilization* cluster are statements that highlight some of the affordances that digitalization can bring. Those are "...greater flexibility for students by recording all lectures in a clear format" (IM = 3.61; FE = 3.77), "...if the digital material like lecture recordings, video tutorials, exercises and scripts can be made more interactive through questions/tests, quizzes, gamification...." (IM = 3.77; FE = 3.63). The cluster also contains statements about requirements for flexibility in the institutional acting. Chief among them is "...the insight: digitalization needs money." (IM = 4.36; FE = 3.91).

*Institutional Ecosystem* contains statements related to the existing e-learning systems and the functionalities that should be available in those. Features that are mentioned are "... Simple search functions in all systems with which you can access the desired information" (IM = 3.79; FE = 2.98), a student dashboard to monitor study progress (IM = 4.02; FE = 3.09), a "...a digital archive with all the old teaching materials (exercise sheets, exams, scripts, transcripts) which are available to students as preparation material" (IM = 3.70; FE = 3.67).
A discrepancy with high importance and low feasibility exists for the unification of the various platforms for planning and organizing the studies (IM = 3.81; FE = 2.63). The least important statement is also contained in this cluster: "...use of chatbots in consulting scenarios" (IM = 2.37; FE = 2.86).

*Hard- & Software Support is* concerned with technology that must be purchased, as well as support for the usage of them. Physical hardware that is requested includes PCs, tablets, video production equipment, as well as "modern systems" in general. Notable are *having sufficient power outlets* (FE = 3.44; IM = 4.09) and *reliable technical support for teachers and students* (IM = 3.98; FE = 3.49). A high difference between importance and feasibility is towards the statements "...investing in and maintaining modern hardware and software that makes digital education possible in the first place" (IM = 4.32; FE = 3.21) and "...that sufficient space is available in the [institution] in which the digital learning technologies can be used (e.g. Virtual Reality, and digital exams)" (IM = 4.00; FE = 2.51).

| Cluster Name | n | BR | m(IM) | $\sigma$(IM) | m(FE) | $\sigma$(FE) |
|---|---|---|---|---|---|---|
| Digital Governance | 8 | 0.68 | 3.70 | 0.08 | 3.37 | 0.16 |
| Cross-Linking of the IT Infrastructure | 8 | 0.40 | 3.06 | 0.27 | 3.27 | 0.14 |
| Flexibilization | 10 | 0.64 | 3.58 | 0.09 | 3.44 | 0.09 |
| Central Data Repositories | 11 | 0.36 | 3.60 | 0.01 | 3.30 | 0.10 |
| Novel Services | 11 | 0.38 | 3.34 | 0.18 | 3.50 | 0.29 |
| Hard- & Software Support | 17 | 0.08 | 3.62 | 0.18 | 3.05 | 0.13 |
| Institutional Ecosystem | 12 | 0.33 | 3.45 | 0.20 | 3.21 | 0.18 |
| Internal and External Cooperation | 13 | 0.82 | 3.48 | 0.09 | 3.49 | 0.12 |
| Critical Accompanying Research | 23 | 0.69 | 3.55 | 0.14 | 3.73 | 0.06 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Cultural Change | 22 | 0.31 | 3.78 | 0.06 | 3.53 | 0.13 |

Tab. 3: Description of the clusters with n = number of participants; BR = Average Bridging value of the statements in the cluster; IM = Average importance rating; FE = Average feasibility rating; $\sigma(\text{IM}) = \text{Variance of the importance rating}$; $\sigma(\text{FE}) = \text{Variance of the feasibility rating}$

### 3.2 Ratings

For the majority of clusters, the importance of the statements is ranked higher than their feasibility. The biggest ($\Delta = 0.57$) and most significant ($p < 0.001$) difference is in the cluster *Hard- & Software Support* where several statements are rated with higher importance than feasibility. The two clusters *Cultural Change* ($\Delta = 0.25$; $p < 0.02$) and *Central Data Repositories* ($\Delta = 0.30$; $p < 0.01$) clusters also have a difference which is statistically significant. The opposite trend, a higher feasibility and a lower importance is observable in the clusters *Novel Services* ($\Delta = 0.25$), *Critical Accompanying Research* ($\Delta = 0.25$) and *Cross-linking of the IT infrastructure* ($\Delta = 0.20$), although in all three cases the results are not statistically significant.

We also investigated differences between the groups that we identified via the questionnaire (see table 1 and 2). First, we compared students and employees. For the group of employees, we combined researchers, management and teachers. We found significant differences in the *Cultural Change* cluster, both regarding feasibility and importance. Employees ranked the importance of cultural change on average higher than the students (IM(Employees) = 3.92; IM(Students) = 3.64; $p < 0.005$), while students ranked the feasibility of cultural change on average higher than the employees (FE(Employees) = 3.43; FE(Students) = 3.70; $p < 0.02$).
Next, we compared novices with less than three years of experience, with experienced users who have more than three years of experience using learning technologies. We found no significant differences between those groups for any of the clusters.

| Cluster | Student | | Employee | | Novice | | Expert | |
|---|---|---|---|---|---|---|---|---|
| | IM | FE | IM | FE | IM | FE | IM | FE |
| Digital Governance | 3.61 | 3.45 | 3.80 | 3.37 | 3.64 | 3.36 | 3.70 | 3.37 |
| Cross-Linking of the IT Infrastructure | 3.11 | 3.28 | 3.12 | 3.32 | 3.08 | 3.18 | 3.05 | 3.33 |
| Flexibilization | 3.48 | 3.42 | 3.62 | 3.46 | 3.67 | 3.38 | 3.51 | 3.49 |
| Central Data Repositories | 3.52 | 3.17 | 3.65 | 3.34 | 3.57 | 3.24 | 3.62 | 3.34 |
| Novel Services | 3.38 | 3.53 | 3.31 | 3.48 | 3.40 | 3.39 | 3.29 | 3.58 |
| Hard- & Software | 3.58 | 3.09 | 3.78 | 3.08 | 3.51 | 2.89 | 3.72 | 3.17 |

| Support | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Institutional Ecosystem | 3.39 | 3.22 | 3.56 | 3.28 | 3.47 | 3.01 | 3.44 | 3.36 |
| Internal and External Cooperation | 3.41 | 3.43 | 3.58 | 3.56 | 3.33 | 3.36 | 3.60 | 3.59 |
| Critical Accompanying Research | 3.53 | 3.73 | 3.60 | 3.75 | 3.48 | 3.68 | 3.60 | 3.78 |
| Cultural Change | 3.64 | 3.70 | 3.92 | 3.43 | 3.80 | 3.54 | 3.77 | 3.53 |

Tab. 4: Average importance (IM) and average feasibility (FE) ratings of the clusters by participant group

## 4 Discussion and Conclusion

The participants of this GCM study have provided many statements and we could observe interesting relations between them. Across all groups of stakeholders, changes in attitude and mentality are rated as the most important statements. Openness towards digitalization, not blindly following trends and cooperation between departments are requested repeatedly. At the same time, many of these statements are also rated with low feasibility relative to their importance. This discrepancy is highest among the group of employees. One can assume that this group has experience with the difficulties involved in achieving cultural changes. This large concern with matters of attitude and mentality could be indicative of a process that is in its earlier stages. In comparison with similar studies from other countries, e.g. the Netherlands or Australia, it is noticeable that these studies discuss more issues of concrete implementation and optimization. In fact, the results that we have obtained are quite comparable with those from studies that are now more than a decade old [Wo05].

On the other hand, topics beyond attitude and mentality, such as wishes for specific hardware or services do exist. They are, on average, seen as less important. Yet this does not necessarily mean that they should not be part of a successful digitalization strategy. The higher variance for these statements shows that these specific wishes are important for some but may not matter for others. On the topic of purchases, it is also notable that statements which require significant investments to be realized, i.e. a “modern” infrastructure, are mostly seen as not very feasible. This again indicates that there are certain experiences and frustrations in place.

What we found surprising but at the same time very encouraging is that there are few differences between the students and the employees for almost all issues.

Those that aim to increase digitalization at higher education institutions should keep these results in mind when planning their strategy. The mentality and attitudes towards digital transformation efforts should be regarded as at least as important as the implementation of new technologies and services. The other way around, simply implementing the technologies without regard to the mentalities, may be less fruitful and it is conceivable that its adoption may be hindered by a lack of acceptance.

While we think that the GCM methodology is suited to capture a representative range of opinions, there are bound to be opinions which were not captured, thus we cannot claim to provide a complete overview.
To summarize the observations, there are still a lot of cultural and policy issues which seem to stand in the way of adoption. Addressing them should be the main focus of digitalization efforts in higher education.

## Bibliography

[Co16] Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., Nelson, K., Alexander, S., Lockyer, L., Kennedy, G., Corrin, L., & Fisher, J. (2016). Student retention and learning analytics: a snapshot of Australian practices and a framework for advancement. Canberra, ACT: Australian Government Office for Learning and Teaching

[Dr14] Drachsler, H., Stoyanov, S., and Specht, Marcus: The Impact of Learning Analytics on the Dutch Education System. In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge; 2014

[IS16] Ifenthaler, D., Schumacher, C.: Student perceptions of privacy principles for Learning Analytics. In Education Technology Research and Development; 2016

[RK12] Rosas, S., Kane, M.: Quality and rigoer of the concept mapping methodology: a pooled study analysis. In Evaluation and Program Planning. Vol. 35, Issue 2, p. 236 – 245, 2012

[Sc17] Schmid, U. et al: Monitor Digitale Bildung. Die Hochschulen im digitalen Zeitalter. Bertelsmann-Stiftung. 2017

[SI18] Schumacher, C., Ifenthaler, D.: Features students really expect from learning analytics. In Computer in Human Behavior. Volume 78, p. 397-407. 2018

[TM17] Trepse, S. Masur, P.K: Privacy attitudes, perceptions, and behaviors of the German population: Research Report. In Friedewald et al. (Eds.), Forum privatheit und selbstbestimmtes Leben in der digitalen Welt; 2017

[Tr89] Trochim, W.M.K. An introduction to concept mapping for planning and evaluation. Evaluation and Program Planning 12, 1 (1989), 1–16.

[Wo05] Wopereis, I. et al: Failure and success factors of educational ICT projects: a group concept mapping approach. In British Journal of Educational Technology. Volume 36, Issue 4, p. 681 – 684. 2005

# Effects of Digitalization on Vocational Education and Training: First Results of a Qualitative Study

Julian Busse[1], Aline Lange[1] and Matthias Schumann[1]

**Abstract:** Digitalization confronts vocational education and training (VET) with both challenges and potentials. However, concrete effects of digitalization are missing in research. Especially the perspective of training and learning departments seems to be interesting because the employees in these areas are the central actors in imparting the necessary competencies and skills to the employees of enterprises. Therefore, the main goal of this paper is to elaborate effects of digitalization on VET from the above-mentioned perspective. To reach this goal, we conducted a qualitative study by interviewing 20 experts of enterprise training and identified two main effects on VET.

**Keywords:** Digitalization, VET, E-Learning, Enterprises, Digital Learning

## 1 Digitalization in VET

For enterprises and thus for vocational education and training (VET), the ongoing digitalization has both, potentials and challenges [DOB19]. On the one side, the digitalization changes job profiles and job structures [Be15], [LLA18], [FO13]. Therefore, work tasks become more complex and more demanding for employees [PN14], [DM15], [HS16]. On the other side, the digitalization offers new potentials for education and learning in enterprises by enabling new forms of technology-based learning like mobile learning or game-based learning [AR16]. Despite this, there is a lack of research that deals with concrete effects of digitalization on VET [Ha18]. Initial empirical findings show, for example, the effects of digitalization in VET on learning content as well as on the design of curricula from the perspective of experts in the fields education and economics [SMM17]. Nevertheless, empirical findings from the perspective of training and learning departments of the enterprises are missing. However, this perspective is particularly relevant because the employees in this area are the central actors in imparting the necessary competencies and skills to the employees of enterprises. Therefore, the main goal of this paper is to elaborate effects of digitalization on VET from the above-mentioned perspective. Thus, we present insights from a qualitative, explorative study to address the following research question:

**RQ1:** What are the effects of digitalization on VET?

In the next section, we present the research method and framing of our study. Afterward, we describe first results of our analysis, before we finally give a conclusion and outlook.

[1] Universität Göttingen, Professur für Anwendungssysteme und E-Business, Platz der Göttinger Sieben 5, 37073 Göttingen, {julian.busse|aline.lange|mschuma1}@uni-goettingen.de

doi:10.18420/delfi2019_206

## 2 Method

To identify what effects digitalization has on VET, we chose to conduct a qualitative and explorative interview study [My13]. Our methodical approach had three steps:

**First**, we identified potential experts from enterprises in different sectors through research in practice-relevant journals. We searched for experts, who work in the field of VET and are in charge of the design of learning content in their enterprises. In order to ensure heterogeneity and to achieve a comprehensive cross-section for the research area, we involved experts independent of the industry. However, a bias for the Banking industry can be derived. We initially contacted the experts via e-mail and sent them a leaflet about the research project. In total, 20 of the potential experts from enterprises in Germany, Austria, and Switzerland accepted our interview invitation. The summarized characteristics of the sample are shown in Tab. 1.

| Expert (Enterprise) | Industry | Expert (Enterprise) | Industry |
|---|---|---|---|
| **Exp1 (U1)** | Customer-Services | **Exp11 (U9)** | Banking |
| **Exp2 (U2)** | Telecommunications | **Exp12 (U10)** | Consulting |
| **Exp3 (U2)** | Telecommunications | **Exp13 (U11)** | Furniture |
| **Exp4 (U3)** | Engineering | **Exp14 (U12)** | Engineering |
| **Exp5 (U4)** | Banking | **Exp15 (U13)** | Automotive |
| **Exp6 (U4)** | Banking | **Exp16 (U14)** | Food |
| **Exp7 (U5)** | Engineering | **Exp17 (U15)** | Banking |
| **Exp8 (U6)** | Chemistry | **Exp18 (U16)** | Banking |
| **Exp9 (U7)** | Aerospace | **Exp19 (U16)** | Banking |
| **Exp10 (U8)** | Banking | **Exp20 (U17)** | Engineering |

Tab. 1: Sample Characteristics

At the **second** step, we conducted all interviews via phone from March to Mai 2018. The interviews lasted between 18 and 80 minutes (Ø= 43 min; $\bar{x}$ =39 min). For conducting the interviews, we used a semi-structured interview guideline where one section of questions addressed the effects of digitalization on VET (e g. *The effects of digitalization include, for example, the increasing automation of routine-based activities, the increase in individualized work steps or the increase in complex, cross-process activities. What effects will these developments have on VET?).* Due to the flexibility of a guideline, the order and the concrete formulation of the questions could be adapted ad hoc during the conversation [Fl14]. At the **third** step, we recorded the interviews on tape and transcribed them afterward, so a further in-depth-analysis was possible. With the help of a structured content analysis approach, we analyzed the empirical data [Ma14]. In this step, we anonymized the transcribed recordings and coded the relevant statements by using the software MAXQDA. The coding was conducted by two independent researchers through a continuous analysis of the transcripts, followed by a mapping of the codes to the core topics. Afterward, we translated the relevant statements into English.

## 3 Findings

In order to answer RQ1, we identified two main effects of digitalization on VET - Organization of VET (OVET, see Tab. 2) and Increasing Learning Needs (ILN, see Tab. 3). Regarding **OVET**, we identified three categories that describe changes because of increasing digitalization. Six experts stated that the effects of digitalization change the way of learning in enterprises from traditional forms like classroom training to various forms of e-learning (**OVET1**).

| OVET | Exemplary Quotes |
|---|---|
| **OVET1: Shift from traditional forms of learning to e-learning (n=6)** | "We have a large number of employees and topics that affect many employees, which, if the learning content allows it, can be offered more easily via digital learning or at least make supplementary offers. This has grown very strongly in recent years." (Exp15) |
| | "We are doing more webinars instead of seminars in order to be able to present content at short notice without having to bring the whole staff into the head office." (Exp10) |
| **OVET2: Enrichment of learning opportunities (n=5)** | "Digitalization offers many more opportunities for learning because the workforce is also distributed with smartphones, tablets or laptops, which can be used to access e-learning. [...] there are many more opportunities for VET." (Exp20) |
| **OVET3: Individualization of educational pathways (n=3)** | "The individualization of education pathways is becoming more and more exciting. [...]. For example, we no longer have a standard training catalog, [...] instead it is much more the approach: "I have to pick you up individually where you are (according to individual learning needs)"." (Exp13) |

Tab. 2: Organization of VET

Nevertheless, the experts made it clear that e-learning does not replace classical forms of learning but is much more of a complementary learning offer. E-learning enables enterprises to reach a large number of employees with less effort than classroom training. Especially for topics that affect many employees, the experts consider e-learning as a very useful option to convey the necessary knowledge on a large scale. A central driver in this context is an increasing need for place and time independent learning. Therefore, the place of learning is not a static dimension anymore and getting more variable. Thus, learning concepts like mobile and micro learning increasingly obtain more importance because they can address this development and enable, for example, work integrated and situated learning. The next effect, also related to a shift to e-learning, is an **enrichment of learning opportunities** in VET (**OVET2**). A central driver that makes this possible is an increase distribution of mobile devices like smartphones, tablets or laptops among the workforce. These represent the necessary access technology to be able to reach employees with various forms of e-learning. Thus, enterprises have the opportunity to enrich their learning offers and, for example, offer situation-related learning content with technology-enhanced

forms of learning, e. g. in the form of short learning units (micro learning) on mobile device (mobile learning). Furthermore, the widespread use of e-learning makes it possible for learning to take place in a less time-bound manner. The learning process can thus be extended depending on the needs so that enterprises can increasingly promote self-determined learning of their employees. Along with that, an **individualization of educational pathways** takes place (**OVET3**). The learning opportunities offered by enterprises are increasingly changing from training measures, which have to appeal to a broad audience, to individual education and training paths. This development is supported in the context of digitalization by the fact that data on the learning progress of learners can be analyzed in a simpler way (learning analytics). Thus, new and individual learning paths become possible which can increase the effectiveness of the training measures.

Furthermore, the experts see an **increase of learning needs (ILN)** because of digitalization. They describe that working life is becoming more fast-paced. For example, products change at ever-shorter intervals, which at the same time increases the learning needs of employees (e .g. training for products or for new production processes). In this context, the experts observe that the learning frequency increases but the learning duration decreases. In addition, learning content is faster becoming obsolete. Thus, the experts state that it is crucial to integrate learning at the workplace, for example, to reduce interruptions of the actual working process. In the next paragraph, we show three different categories that the experts described in this context.

The biggest change perceived by the experts due to the increased need for learning is the growing **importance of demand-oriented learning** (**ILN1**). According to the experts, this development is closely linked to the concept of performance support. This means that employees spend less time in classic classroom training and receive more support at their moment of need, e. g. at their workplace when a problem with a machine occurs. As digitalization makes the activities of employees more complex and less predictable, this form of learning is becoming increasingly important. In order to support the employees, situational and demand-oriented learning content is necessary. The experts describe that the provision of learning content is, therefore, evolving from a push to a pull mechanism. Thus, employees mainly learn when they consider the learning content to be relevant or when they have a specific problem. Closely linked to this aspect is the **demand for small learning contents**, as these are particularly suitable to satisfy the employees' demands for individual and situation-specific support with technology-enhanced learning formats in their moment of need (**ILN2**). The experts attach great importance to these small and short learning contents, especially since only these characteristics make it possible to reduce work interruptions to a minimum and still integrate learning into every day working life. For example, one expert in the banking branch sees great potential for small and short information units to inform field staff about market reactions or new products via their mobile devices (Exp18). Therefore, the enterprises spend great efforts to keep the (e-)learning content as short as possible. Another important point mentioned by the experts is that a **high level of knowledge** is increasingly becoming a critical competitive factor for the enterprises (**ILN3**). Customers expect their business partners to have a high level of knowledge as digital information platforms are increasingly making the market

situation more transparent and competitors more comparable. This problem is particularly relevant in the retail sector, where customers have a good overview of the market via numerous digital information platforms such as price comparison portals. If the suppliers and business partners do not have significantly better and more up-to-date knowledge as their customers, the customers may perceive the advice as superfluous [SFV16].

| ILN | Exemplary Quotes |
|---|---|
| **ILN1: Demand-oriented learning and performance support (n=9)** | "Due to the diversity of the topics and also due to the technical possibilities it changes in such a way that more and more is learned at the moment of need." (Exp14) |
| | "So if someone learns for himself in 70% of the cases [70:20:10 model] simply because he or she is personally faced with a task, then I have to rather see what is so possible on-demand. Digitalization also helps here." (Exp13) |
| **ILN2: Need for short learning units (n=3)** | "The contents must be short so that the employees are not delayed for long in the normal workflow." (Exp7) |
| | "In the last three years we have shortened our e-learning by more than half and now offer learners formats of less than a minute." (Exp3) |
| **ILN3: High level of knowledge as a competitive factor (n=2)** | "Learning [...] is becoming more and more central. [...]. The customer has high demands on his contact person and his supplier and there must be a very high level of knowledge." (Exp4) |

Tab. 3: Increasing learning needs

## 4 Conclusion and Outlook

Bringing it together, our findings show a positive perspective on the effects of digitalization on VET from the view of training and learning departments. The technological possibilities complement classical measures of VET and allow new learning opportunities like performance support at the workplace with micro content on mobile devices. E-Learning helps enterprises to communicate important information as well as learning content to their employees in a faster and more efficient way. This also makes it possible to design educational offers more individually and to tailor them better to the learner with the help of data analysis. In addition, the experts stated that the new e-learning formats help to shift the learning more to the place of the event or the moment of need. Thus, employees receive information in their moment of need, for example at their workplace or on-site at the customer premises. However, what remains unconsidered by the experts are the problems and challenges that enterprises perceive with regard to digitalization. Since the successful implementation of e-learning depends to a large extent on the appropriate didactical preparation of the learning content, future research should

examine which problems and challenges enterprises perceive in the didactical preparation of learning content. These findings would make a major contribution to both science and practice in making digital learning more successful in enterprises as well as in other contexts.

## Bibliography

[AR16] Alkhatib, W.; Rensing, C.: Towards a Classification of Learning Support Systems at the Digitized Workplace. In: Proceedings of DeLFI Workshops 2016, Potsdam, 2016, pp. 188-194.

[Be15] Bertschek, I.: Industrie 4.0: Digitale Wirtschaft. Herausforderung und Chance für Unternehmen und Arbeitswelt. ifo Schnelldienst 10, pp. 3-5, 2015.

[DM15] Dengler, K.; Matthes, B.: Folgen der Digitalisierung für die Arbeitswelt. Substituierbarkeitspotenziale von Berufen in Deutschland, Nürnberg, 2015.

[DOB19] Distel, B.; Ogonek, N.; Becker, J.: eGovernment Competences revisited – A Literature Review on necessary Competences in a Digitalized Public Sector. In: Proceedings of 14th International Conference on Wirtschaftsinformatik, Siegen, 2019, pp. 296-310.

[Fl14] Flick, U.: An introduction to qualitative research. Sage, Los Angeles, 2014.

[FO13] Frey, C. B.; Osborne, M. A.: The future of employment: how susceptible are jobs to computerization? Oxford University, Oxford, 2013.

[Ha18] Harteis, C.: Machines, Change and Work: An Educational View on the Digitalization of Work. In: The Impact of Digitalization in the Workplace. An Educational View. Springer, Cham, pp. 1-10, 2018.

[HS16] Hammermann, A.; Stettes, O.: Qualifikationsbedarf und Qualifizierung. Anforderungen im Zeichen der Digitalisierung. IW policy paper, No. 3/2016, Köln, 2016.

[LLA18] Lindner, D.; Ludwig, T.; Amberg, M.: Arbeit 4.0 - Konzepte für eine neue Arbeitsgestaltung in KMU. HMD Praxis der Wirtschaftsinformatik 5/18, pp. 1065-1085, 2018.

[Ma14] Mayring, P.: Qualitative Content Analysis. Theoretical Foundation, Basic Procedures and Software Solution. Beltz, Klagenfurt, 2014.

[My13] Myers, M. D.: Qualitative research in business & management. Sage, London, 2013.

[PN14] Picot, A.; Neuburger, R.: Arbeit in der digitalen Welt. Zusammenfassung der Ergebnisse der AG 1-Projektgruppe anlässlich der IT-Gipfelprozesse 2013 und 2014, München, 2014.

[SFV16] Schildhauer, T.; Flum, T.; Voss, H.: Weiterbildung im Kontext der Wirtschaft 4.0. Controlling 4-5, pp. 266-272, 2016.

[SMM17] Sachs, S.; Meier, C.; McSorley, V.: Digitalisierung und die Zukunft kaufmännischer Berufsbilder – eine explorative Studie. https://bit.ly/2IodmDF, accessed: 12/02/2019.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 73

# Weißt Du es oder rätst Du noch? – ARS mit Rückmeldungen zur Antwortzuversicht bei Lernfragen

**Prototypischer Vorschlag und erste Erkenntnisse**

Lucas Braeschke[1], Iris Braun[2], Felix Kapp[3] und Tenshi Hara[4]

**Abstract:** Die reine Erfassung von Antworten auf Lernfragen in Audience Response Systemen (ARS) leidet in der Regel unter einer Entkopplung zwischen tatsächlich vorhandenem Wissen, gefühltem Wissen und irrtümlich angenommenem Wissen. Für Lernende ist eine Einschätzung der Wissensqualität für erfolgreiches Lernen jedoch unverzichtbar. Durch die zusätzliche Erfassung der Antwortzuversicht wird Lernenden der tatsächliche Wissensstand zugänglich; darüber hinaus ist dann eine erfolgreiche Lernbedarfsermittlung möglich. Ob fehlendes oder falsches Verständnis vorliegt, oder Wissen nur partiell oder doch vollständig verarbeitet wurde, kann mit den Mitteln von Audience Response Systemen einfach und elegant verarbeitet und den Lernenden als individualisierte Auswertung zugänglich gemacht werden. In diesem Beitrag wird eine prototypische Implementierung mit ersten Evaluationsergebnissen präsentiert.

**Keywords:** Audience Response System, ARS, Zuversicht, Selbsteinschätzung

## 1 Einleitung

Audience Response Systeme (ARS) bieten die Möglichkeit, Lernende beim Wissenserwerb zu unterstützen. Durch Lernaufgaben, Umfragen und Abstimmungen lässt sich die Interaktivität in der Vorlesung erhöhen. Studierende haben die Möglichkeit, sich aktiv mit relevanten Konzepten auseinanderzusetzen und bei der Bearbeitung von Lernaufgaben auch Rückmeldung hinsichtlich ihres Lernfortschritts zu erhalten. Diese Information unterstützt sie beim erfolgreichen Regulieren ihres Lernprozesses: sie bekommen mit Hilfe von ARS Hinweise, welche Wissensgebiete bereits in der Lehrveranstaltung verstanden wurden und welche im Nachgang nochmals thematisiert werden sollten. Die häufig verwendeten Multiple-Choice-Fragen fordern die Lernenden dazu auf, die korrekte Antwort auf die Lernfrage auszuwählen, wobei das Potenzial dieser Lernaufgaben darin liegt, den Lernprozess sowohl in der aktiven Auseinandersetzung mit dem Inhalt während der Bearbeitung, als auch im sich anschließenden Reflexionsprozess, der durch die Beantwortung angestoßen werden kann, zu unterstützen [Ka11]. Die Lernenden

[1] Saxonia Systems AG, Fritz-Förster-Platz 2, 01069 Dresden, lucas.braeschke@saxsys.de
[2] Professur für Rechnernetze, Fakultät Informatik, TU Dresden, 01062 Dresden, iris.braun@tu-dresden.de
[3] Institut für Psychologie und Arbeitswissenschaft, FG Mensch-Maschine-Systeme, TU Berlin, Fasanenstr. 1, 10623 Berlin, felix.kapp@tu-berlin.de
[4] Berufsakademie Sachsen – Staatliche Studienakademie Dresden, Hans-Grundig-Str. 25, 01307 Dresden, tenshi.hara@ba-dresden.de

doi:10.18420/delfi2019_338

können die Rückmeldung verwenden, um ihren Lernprozess neu auszurichten, ihre Konzentration, Lernstrategien und Lernziele auf die aktuellen Gegebenheiten anzupassen. Wird neben der eigentlichen Multiple-Choice-Lernaufgabe noch zusätzlich die Antwortzuversicht abgefragt, kann dieser Vorgang effektiver gestaltet werden. Das Ziel ist es, besser auf Wissenslücken aufmerksam zu machen und somit auch den Lernprozess zu verbessern. So soll der Unterschied zwischen der eigenen Selbsteinschätzung und dem tatsächlich Gelernten aufgezeigt werden. Vor allem sind die Fälle, in denen diese beiden Faktoren divergieren, als besonders problematisch zu betrachten. Etwa wenn eine Frage trotz einer besonders hohen Zuversicht falsch beantwortet wurde oder aber wenn keine Zuversicht in die eigene richtige Antwort bestand. Diese Informationen sind ebenfalls für Lehrende wichtig. Nur wenn sie von etwaigen Problemen wissen, können sie diese auch adressieren. Vor diesem Hintergrund erscheint das Potenzial von Lernaufgaben ergänzende Zuversichtserfassung für die universitäre Lehre groß. Dabei können die erhobenen Daten sowohl unmittelbar in der Lehrveranstaltung zur Reflektion anregen und die Lehrenden unterstützen als auch zu einem späteren Zeitpunkt bei der Prüfungsvorbereitung und der Auswahl von Lernstoff und -strategien helfen.

Mit Auditorium Mobile Classroom Service (AMCS)[5] werden kontinuierlich didaktische Konzepte mit den technischen Mitteln von ARS zusammengeführt. AMCS unterstützt dabei unterschiedliche Feedback-Mechanismen, u.a. auch Multiple-Choice-Fragen. Die Antworten auf die Fragen können als Datenbasis für die Lehrveranstaltungsevaluation, Lernstandanalysen sowie das Auslösen kognitiver und meta-kognitiver Prompts verwendet werden [Ka14]. Im Rahmen dieses Beitrags wird gezeigt, wie die zuvor genannte Zuversichtserfassung erfolgreich in ARS eingebracht und mit anschließender Zuversichtsbewertung Hinweise für die Prüfungsvorbereitung gegeben werden können.

## 2 Verwandte Arbeiten

Zuversicht (engl. *Confidence*) beschreibt laut Duden das „feste Vertrauen auf eine positive Entwicklung in der Zukunft“. Teilweise, wenn auch weniger verbreitet, wird in diesem Zusammenhang auch Sicherheit/Bestimmtheit (engl. *Certainty*) genutzt. Im Kontext der Lehre kann eine Frage nach der Zuversicht eingesetzt werden, um zu überprüfen, wie sicher sich die Teilnehmer bei der Beantwortung einer Frage sind.

Um die Zuversicht der Teilnehmer bei einer Lernfrage zu messen, wird meist eine vierstufige Skale genutzt (z.B.: „keine“, „geringe“, „mäßige“ und „hohe“ Zuversicht [Cu13],[LB16]). Die häufigste Anwendung findet die Zuversichtserfassung bei *Multiple-Choice-Fragen* (MCQ). Es existieren sowohl Varianten, bei denen die Sicherheit abgefragt wird, bevor die Antwortoptionen angezeigt werden (*Open Confidence*), als auch jene bei denen erst nach der Auswahl einer Antwort die Zuversicht erfasst wird (*Certainty-based Marking*).

[5] https://amcs.website/

## 2.1 Zuversichtserfassung in Audience Response Systemen

In Audience Response Systemen wird eine Zuversichtserfassung bisher nur sehr selten umgesetzt. In einer Evaluation von 50 ARS und deren Funktionen [Ku19] wurde nur ein ARS identifiziert, das bisher Zuversichtserfassung umsetzt. Das an der TH Mittelhessen entwickelte *ARSnova* Projekt enthält in seiner vereinfachten Version arsnova.click[6] auch die Abfrage der Zuversicht bei der Beantwortung von Lernaufgaben. Das System unterstützt fünf verschiedene Frage-Typen, wobei es sich bei vier um Lernfragen handelt. Die Zuversicht kann mittels einer Skala ausgewählt werden, welche in fünf Abstufungen unterteilt ist. Die Ausprägung auf der Skala ist neben der Beschriftung auch farbkodiert: *geraten* (dunkelrot), *sehr unsicher* (hellrot), *unsicher* (ocker), *relativ sicher* (hellgrün), *absolut sicher* (dunkelgrün). Die Antwortoption(en) und die Ausprägung der Zuversicht werden zeitgleich durch eine Schaltfläche abschließend bestätigt. Die Erfassung der Zuversicht ist sehr eingängig umgesetzt und lässt sich leicht bedienen und verstehen.

Wenn die Teilnehmenden die Frage beantwortet haben, wird für alle Nutzer eine Gruppenauswertung angezeigt, welche die gewählten Antworten und die durchschnittliche Zuversicht darstellt. Darüber hinaus beinhaltet die Gruppenauswertung ein Ranking der Teilnehmenden auf der Basis der Antwortzeiten. Dieses *Gamification*-Element konterkariert allerdings das didaktische Ziel der Zuversichtserfassung, nämlich Studierende dazu anzuregen, zu reflektieren, welchen Wissensstand sie in einem Wissensgebiet haben. Da schnelles Antworten im Ranking belohnt wird, ist eine tiefergehende Auseinandersetzung mit dem eigenen Wissenstand eher unwahrscheinlich.

## 2.2 Bestimmtheitsbasierte Markierung (Certainty-based Marking)

*Certainty-Based Marking* (CBM) ist in der Literatur auch unter dem Namen „*London Agreed Protocol for Teaching*“ bekannt. Entwickelt wird es seit 1994 am University College London (UCL). Lernende können dabei nicht nur mit einer Vielzahl an Testfragen Selbsttests durchführen, sie bekommen auch eine Rückmeldung inwiefern sie sich bereits präzise selbsteinschätzen können. In der webbasierten Umsetzung[7] der UCL finden sich über 7000 Fragen. Den wichtigsten Teilbereich macht dabei der „Medical and Biomedical Students' Self-Test“ mit über 5000 Einträgen aus.

Die Angabe der Zuversicht („*unsure*“, „*mid*“, „*sure*“, „*no idea*“ (Enthaltung)) erfolgt beim CBM zeitgleich zur Auswahl der Antworten auf die Lernfragen. Die Auswertung für die Lernenden beinhaltet eine Leistungskomponente (als *Präzision* bezeichnet; durchschnittlich erreichte Punkte) sowie eine Übersicht der Selbsteinschätzung (wie zuversichtlich die Studierenden bei der Beantwortung waren). Die Kombination beider Komponenten erlaubt eine Aussage über die Genauigkeit der Selbsteinschätzung der Lernenden zu machen. Dafür fließt die Zuversicht über den *CBM-Bonus* ein, welcher in Relation zum durchschnittlichen CBM-Wert für die entsprechende Präzision berechnet

[6] https://arsnova.click/
[7] http://www.ucl.ac.uk/lapt/

wird. Der Bonus wird vergeben, wenn der erzielte CBM-Wert nicht dem erwarteten durchschnittlichen Präzisionswert entspricht; er soll die abweichende Selbsteinschätzung in die Präzision einbeziehen. Dies gilt ebenfalls für den negativen Bereich, also wenn der eingeschätzte Wissensstand deutlich höher ist, als dies die Ergebnisse vorweisen. Der CBM-Bonus wird mit der Präzision zu einer allgemeinen Genauigkeit verrechnet. Die eigenen Ergebnisse im Kontext der Antworten anderer Nutzer zu sehen, bietet den Lernenden zusätzliche Möglichkeiten, sich selbst besser einzuschätzen.

### 2.3 Zuversichtsbasierte Markierung (Open Confidence-based Marking)

*The Open University*[8] hat einen interessanten Ansatz im Kontext der Zuversichtseinschätzungen entwickelt, welcher online[9] getestet werden kann. Die entwickelte Methode „*Open Confidence-based Marking*“ (OCBM) ist eine Variante des CBM. Der Unterschied beider Methoden liegt in dem Zeitpunkt der Einschätzung der Zuversicht. Ist bei CBM die Zuversicht erfasst worden, nachdem die Antwortoptionen eingesehen und ausgewählt wurden, wird dies bei OCBM durchgeführt, bevor diese angezeigt werden. Die Einschätzung der Zuversicht besteht wie bei der Umsetzung von UCL aus einer dreistufigen Skala (*„low“, „medium“, „high“*), außerdem gibt es noch die Möglichkeit, mit „*give up*“ die Frage nicht zu beantworten. Zunächst sind nur der Fragetext und das Aufklappmenü für die Zuversicht zu sehen. Erst nachdem die Zuversicht ausgewählt wurde, werden die Antwortoptionen angezeigt. Nachdem eine Antwortoption ausgewählt wurde, wird das Feedback angezeigt. Darunter folgt eine Erläuterung zur richtigen Antwortoption. In der Auswertung wird für jede beantwortete Frage angezeigt, wie viele von den maximal zu erreichenden Punkten nach CBM-Schema erzielt wurden.

Durch die fehlenden Antwortmöglichkeiten zum Zeitpunkt der Selbsteinschätzung bezieht sich die Zuversichtserfassung auf ein größeres Wissensgebiet bzw. ist von generellerer Natur. Im Gegensatz zu Multiple-Choice-Aufgaben, bei denen man vorgegebene Antworten wiedererkennen oder ausschließen kann, beziehen sich die Einschätzungen hier auf Freitextaufgaben (da zum Zeitpunkt der Bewertung noch keine Antwortmöglichkeiten dargestellt werden). Es ist zu vermuten, dass die Aufgaben zu diesem Zeitpunkt als schwerer wahrgenommen werden und somit die Studierenden in ihrer Einschätzung auch unsicherer sind.

## 3 Erfassung und Bewertung der Zuversicht in AMCS

Wie in allen Audience Response Systemen bilden Live-Befragungen auch in AMCS die Kernfunktionalität [Ka14]. Jede Befragung besteht aus einer oder mehreren Lernaufgaben oder Umfragen, die die Studierenden über ihre eigenen Endgeräte (*Bring your own Device*) in Echtzeit beantworten können. Dabei wird zwischen folgenden Typen von Fra-

[8] http://www.open.ac.uk/
[9] https://students.open.ac.uk/openma rk/t216-11b.demo.v03/

gen unterschieden: Umfragen mit Einfachauswahl (*Single Choice*), Umfragen mit Mehrfachauswahl (*Multiple Choice*), Lernaufgaben mit Mehrfachauswahl und einer richtigen Antwort (*Single-Best Choice*) oder mehreren richtigen Antworten (*Multiple-Best Choice*), Freitext-Aufgaben (*Freetext*), Zuordnungsfragen (*Correct Assignment*), sowie Skalenfragen (*Scaled Questions*).

Im Falle von Lernaufgaben und Zuordnungsfragen erhalten die Studierenden direkt nach der Beantwortung neben der Information, ob die Frage richtig oder falsch beantwortet wurde, ein textuelles Feedback mit Hinweisen, warum einzelne Antworten richtig, nicht vollständig oder gar falsch sind. Im Unterschied zu anderen ARS bietet AMCS hier die zusätzliche Möglichkeit, die Frage ein weiteres Mal zu beantworten. Im Anschluss wird die richtige Lösung angezeigt, indem die falschen und die richtigen Antwortmöglichkeiten unterschiedlich farblich markiert werden. Dieser zweistufige Feedbackalgorithmus wurde mit dem Ziel implementiert, dass die Studierenden zur tieferen kognitiven Auseinandersetzung mit dem jeweiligen Inhalt angeregt werden.

Für die Integration der Zuversichtserfassung wurden diese Lernaufgaben ausgewählt, da hier eine direkte quantitative Auswertung der Ergebnisse und damit des Lernerfolges möglich sind. Somit können die Lernenden eine direkte Rückmeldung erhalten, ob und wieviel ihr Lernerfolg und ihre Selbsteinschätzung voneinander abweichen. Es wurden dabei sowohl Zuversichtserfassungen nach/während der Abgabe der Antworten (*Certainty Rating*) als auch vor der Anzeige der möglichen Antworten (*Open Confidence Rating*) prototypisch in AMCS umgesetzt. Die Auswertung der Aufgaben inkl. der notwendigen Zuversichtsbewertung erfolgte sowohl für einzelne Lernaufgaben als auch als aggregierte Werte über mehrere Aufgaben hinweg.

### 3.1 Lernfragen mit Zuversichtserfassung

Die Erfassung der Zuversicht während oder nach Abgabe der Antwort ermöglicht Lernenden eine genauere Einschätzung ihrer Leistung. Reicht das vorhandene Wissen aus, um die Frage mit hoher Sicherheit richtig zu beantworten, oder kann nur geraten werden? Die meisten der verwandten Arbeiten nutzen eine drei- oder fünfstufige Likert-Skala für die Eingabe der Zuversicht. In einer ersten Nutzerstudie wurden daher verschiedene Skalen für die Erfassung der Zuversicht in AMCS getestet und evaluiert, die Ergebnisse werden in Abschnitt 4 beschrieben. Schlussendlich wurde die Nutzung einer Prozent-Skala von 0 bis 100% ausgewählt, da dies eine genauere Eingabe der Zuversicht ermöglicht. Studierende können dazu einen Schieberegler verwenden, der nach der Frage und den möglichen Antworten eingeblendet wird (zur Umsetzung im AMCS-Prototyp siehe Abb. 1).

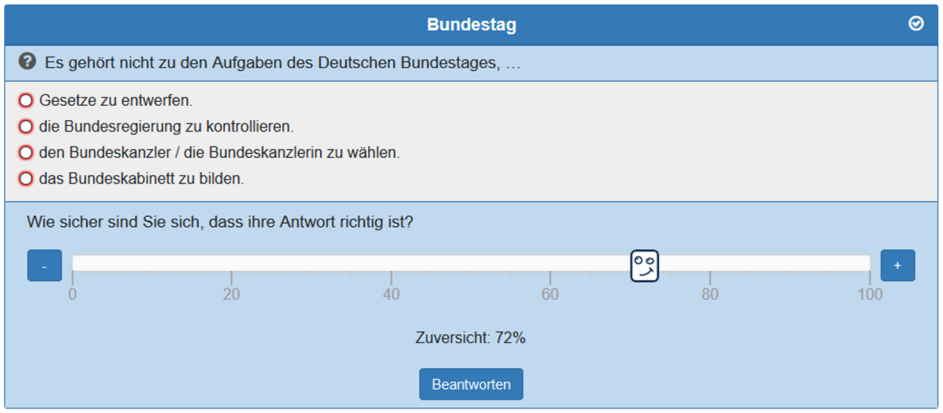



Abb. 1: Multiple-Choice-Frage mit Zuversichtserfassung

### 3.2 Lernfragen mit offener Zuversichtserfassung

Im Gegensatz dazu wird beim *Open Confidence Rating* die Zuversicht bereits vor der Anzeige der möglichen Antworten abgefragt (siehe Abb. 2). Dies erfordert vom Lernenden eine etwas tiefergehende Auseinandersetzung mit der Fragestellung: Kennen sie die Antwort ganz sicher, weil sie das Themengebiet sehr gut verstanden haben, oder wissen sie nur eine ungefähre Antwort? Erst nach Auswahl der Zuversicht gelangen die Lernenden zur eigentlichen Ansicht der Fragestellung mit allen möglichen Antworten.

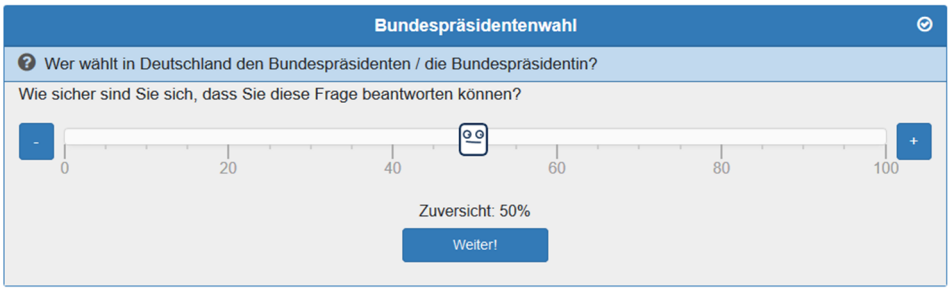



Abb. 2: Abfrage der Zuversicht vor der Beantwortung einer Lernfrage

### 3.3 Zuversichtsbewertung: Evaluation der Zuversichtserfassung

In der Literatur werden Lernende meist in vier Kategorien hinsichtlich der Kombination aus gezeigter Leistung und selbsteingeschätzter Zuversicht eingeteilt (etwa [Hu03], [Cu13]). Man kann dabei zwischen folgenden Kategorien unterscheiden:

**Kein Verständnis (*uninformed*):** In [Cu13] gilt ein Teilnehmer beim Beantworten einer Frage als „*uninformed*“, wenn dieser falsch geantwortet hat und angibt, sich unsicher zu sein. Dies zeigt nach [WG10] einen „*teaching moment*“ an, also eine Möglichkeit und einen Bedarf, mit Hilfe von Feedback und zusätzlichen Erläuterungen auf Lernende einzuwirken und so Wissenserwerb zu fördern.

**Falsches Verständnis (*misinformed*):** Der Status „*misinformed*“ beschreibt Lernende, welche die falsche Antwort ausgewählt haben, jedoch sehr zuversichtlich waren, die richtige gewählt zu haben [Cu13]. Dabei handelt es sich um eine fehlerhafte Selbsteinschätzung, die zu nicht adäquaten Strategien im Vorfeld von Prüfungen führen und das erfolgreiche Meistern dieser gefährden kann.

**Partielles Verständnis (*partial knowledge*):** Partielles Verständnis kann in zwei Fälle unterteilt werden [Bu02, Seite 807]. Grundsätzlich wurde die Frage richtig beantwortet, zugleich jedoch auch angegeben, sich mit der Antwort nicht sicher zu sein. Dies kann zum einen mangelndes Vertrauen in das eigene Wissen bedeuten. Zum anderen könnte es sich auch um den Versuch handeln, die richtige Antwort basierend auf dem lückenhaften Wissen zu erraten. Der erste Fall birgt die Gefahr, dass im Zuge der Prüfungsvorbereitung zu viel Zeit auf ein bereits gelerntes Thema verwendet wird.

**Umfassendes Verständnis (*informed*):** Dieser Fall beschreibt Lernende, welche die richtige Antwort auswählen und gleichzeitig angeben, sich mit der Antwort sicher zu sein. In diesem Fall kann davon ausgegangen werden, dass das entsprechende Wissen erlangt wurde und richtig angewendet werden kann.

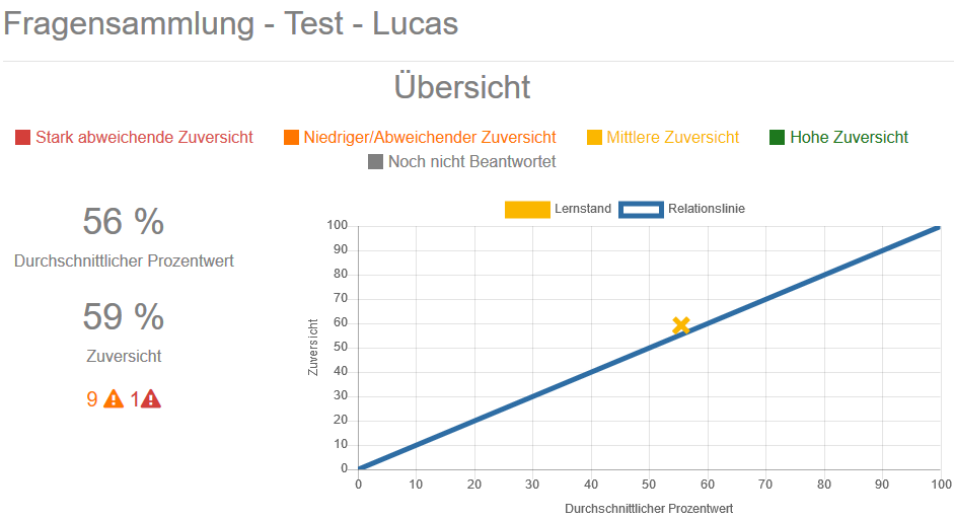



Abb. 3: Auswertung der Zuversicht (Ordinate) im Bezug zum Lernerfolg (Abszisse)

Die Auswertung in AMCS (siehe Abb. 3) hat zum Ziel, die Lernenden bei der Reflektion ihres Lernprozesses zu unterstützen und enthält die gleichen Informationen, welche auch

den vier vorgenannten Typen zu Grunde liegen. Durch eine farbliche Markierung angelehnt an die Ampelfarben [TT14] werden diese Abstufungen der Zuversichtseinschätzung verdeutlicht. Eine adäquate Selbsteinschätzung würde sich als linearer Fit von Zuversicht (Ordinate) und der durchschnittlichen Leistung in Prozent (Abszisse) darstellen (in Abb. 3 als blaue Linie dargestellt). Die eigene Zuversichtsbewertung gemittelt über alle beantworteten Fragen ist dann eine Koordinate mit Abweichung vom Fit, aus der die Lernenden leicht erkennen können, ob sie sich über- oder unterbewertet haben. Die Prozentwerte der Leistung unterscheiden sich für die verschiedenen Fragetypen in ihrer Ausprägung. Für MBC- und CA-Fragen können Werte zwischen 0 und 100% erreicht werden, bei SBC nur 0 oder 100 %. Je mehr Fragen unterschiedlicher Art beantwortet wurden, umso repräsentativer sind die errechneten Werte.

Der beispielhaft dargestellte Lernstand in Abb. 3 ist somit sehr nah am Ideal für leicht über das mittlere Niveau erreichte korrekte Antworten. Die Zuversicht zeigt eine minimale Tendenz zur Selbstüberschätzung. Die Koordinate ist deshalb so nah am Fit, weil beim mittleren Antworterfolg eine dazu passende mittlere Zuversicht vorliegt (in Zahlen: 56% Erfolgsquote bei 58% Zuversicht).

### 3.4 Fragenkatalog

Basierend auf der Zuversichtsbewertung wird ein Fragenkatalog sowohl für die gezielte Wiederholung von Problemfragen (keine oder nicht vollständig richtige Antworten) als auch für die Korrektur von fehlerhafter Wissenseinschätzung (hohe Zuversicht bei falscher Antwort oder niedrige Zuversicht bei richtiger Antwort) erstellt. Idealerweise wird der Fragenkatalog den Lernenden zur Auswahl gestellt, wodurch diese selbst Schwerpunkte festlegen können. Um die Lernenden bei der Auswahl der zu wiederholenden Fragen zu unterstützen, empfiehlt sich eine Markierung der Wiederholungsdringlichkeit. Im AMCS-Prototyp wird entsprechender Handlungsbedarf farblich markiert (siehe Abb. 4). Auch hier wurden in Anlehnung an die Ampel-Metapher Rot und Orange als Signalfarben für erhöhten Handlungsbedarf und Gelb und Grün für weniger Dringlichkeit ausgewählt. Zusätzlich wird die Anzahl der Fragen in der jeweiligen Kategorie mit einem Ausrufezeichen hervorgehoben. Grau markierte Fragen wurden noch nicht beantwortet.

## 4 Nutzerstudien

Um eine möglichst intuitive, den Anforderungen der Lernenden gerechte Umsetzung der Zuversichtserfassung zu erreichen, wurden unterschiedliche, auf Visualisierung und Nutzbarkeit ausgelegte Ansätze untersucht. Dazu wurden Probandenbefragungen mit Papier-Prototypen durchgeführt, auf Basis derer Ergebnisse erste prototypische Änderungen am AMCS-Prototyp vollzogen wurden. Diese wurden dann auf Basis des User Experience Questionnaire[10] (UEQ) vergleichend gegenübergestellt. Zusätzlich wurden

[10] https://www.ueq-online.org

die Probanden bezüglich Farbgestaltung und Symbolauswahl befragt. Insbesondere wurde erfasst, ob und wie sich der Arbeitsfluss verbessern ließ.




Abb. 4: Farblich markierte Vorschläge für Wiederholungsaufgaben entsprechend der gegebenen Antworten und der zugehörigen Zuversichtserklärungen (grün: gute Einschätzung … rot: kritisch)

In einer ersten Studie mit fünf Probanden wurde erfasst, wie die Zuversicht bei der Erfassung visualisiert werden kann. Es wurden drei Papierprototypen vorgeschlagen: Doktorhüte (umso mehr, umso zuversichtlicher), Skala (0 bis 100%; höhere Prozentwerte entsprechen höherer Zuversicht) und Emojis (der Zuversicht entsprechende Gesichtsausdrücke vom Frowny bis zum Smiley).

Anschließend wurden die UEQ-Werte für die drei Prototypen erfasst (siehe Tab. 1), mit der Option weitere Probanden im Falle großer Abweichungen zu befragen. Die Doktorhüte erreichen dabei auf allen Dimensionen schlechtere Werte im Vergleich zu den anderen beiden Visualisierungen, welche wahlweise präferiert werden. Basierend auf den zusätzlich erfassten freien Rückmeldungen der Probanden kann die fehlende Originalität in Kombination mit schlechter Identifikation mit den Doktorhüten als negativer Faktor identifiziert werden.

Den Probanden wurden drei Skalenversionen präsentiert: in 1%-, 10%- und 20%-Schritten. Diesen einfachen Skalen mangelt es eindeutig an Ästhetik. Im Bereich der Neuheit wurden sie sogar negativ bewertet. Dies wiederspricht gängigen Erfahrungen, da Skalen

sonst ein viel verwendetes und daher bekanntes Visualisierungsinstrument in ARS sind und ist eventuell auf die kleine Stichprobengröße zurückzuführen. Die Emojis haben eine hohe Attraktivität, Verlässlichkeit und Neuheit sowie zeichnen sich durch eine deutlich geringere Varianz aus, als die anderen Prototypen. Dies macht sie offensichtlich zur geeignetsten Visualisierung – zumindest für AMCS.

| | **Doktorhüte** | | **Skala** | | **Emojis** | |
|---|---|---|---|---|---|---|
| | **AVG** | **VAR** | **AVG** | **VAR** | **AVG** | **VAR** |
| Attraktivität | 1,3 | 1,41 | 0,83 | 2,33 | 1,93 | 0,43 |
| Klarheit | 1,4 | 1,64 | 1,95 | 0,39 | 1,7 | 1,23 |
| Effizienz | 1,85 | 0,68 | 2 | 0,59 | 1,35 | 1,27 |
| Verlässlichkeit | 1,6 | 1,21 | 1,8 | 0,73 | 1,3 | 0,61 |
| Stimulation | 1,4 | 1,52 | 0,65 | 1,58 | 1,7 | 0,42 |
| Neuheit | 0,85 | 1,02 | -0,05 | 4,08 | 1,8 | 0,98 |

Tab. 1: Auswertung zu den drei Prototypen
(Skalen von -3 – „entsetzlich schlecht“ bis +3 – „extrem gut“; entsprechend UEQ[10])

Nach der Integration eines Emoji-Schiebereglers (siehe Abb. 1 und Abb. 2) wurde eine zweite Befragung durchgeführt. Ziel war es, Probleme bei der Eingabe und dem Arbeitsfluss der Fragenbeantwortung und Zuversichtserfassung zu identifizieren. Dazu wurde der Prototyp der AMCS App in einer echten Lehrveranstaltung eingesetzt und im Anschluss ein UEQ[10] von den 14 Studierenden ausgefüllt. Mit einer durchschnittlichen Klarheit von 2,3[11] ist ein deutliches Indiz für die Eignung dieser Visualisierungsform gegeben. Basierend auf freien Rückmeldungen der Studierenden kann festgestellt werden, dass die Emojis selbsterklärend und einfach zu bedienen sind. Die von uns implementierte Kombination aus Emojis mit Schieberegler[12] erlaubt darüber hinaus eine feingranulare Zuversichtserfassung gegenüber der alleinigen Verwendung von Emojis.

Für die dritte Studie wurde die Erstellung einer Fragensammlung auf Grundlage der Leistungsdaten und Zuversichtsbewertung im Prototyp implementiert. Damit sollte untersucht werden, ob und wie die Vorschläge den Studierenden bei der Prüfungsvorbereitung helfen. Zehn Probanden wurden gebeten sich in eine fiktive Lehrveranstaltung zu versetzen, in welcher sie 20 Fragen zu Politik, Geografie und Geschichte Deutschlands (entnommen aus dem *Einbürgerungstest*) bearbeiteten. Im Anschluss bekamen alle Probanden die Möglichkeit, sich eine Übersicht ihrer Ergebnisse anzuschauen und Fragensammlungen nochmals in Vorbereitung auf eine zukünftige Prüfung zu bearbeiten. Für diese Wiederholungsfragen standen unterschiedliche, vom Prototyp automatisch generierte Fragenkataloge zur Auswahl: 1. alle Lernfragen, unabhängig vom Antworterfolg und/oder Zuversicht, 2. keine Wiederholung, 3. Fragen mit abweichenden Bewertungen (insbesondere für *misinformed* und *partial knowledge*), 4. nicht richtig beantwortete

[11] Die Tabelle zur UEQ-Auswertung fehlt hier aus Platzgründen.
[12] Auf Vorschlag der Probanden aus der ersten Befragung; durch Verschieben des Reglers ändert sich das Emoji in der Schubschaltfläche.

Fragen, 5. Fragen mit Antworten niedriger Zuversicht, 6. Fragen mit Antworten hoher Zuversicht, sowie 7. noch nicht beantwortete Fragen.

In Tab. 2 sind exemplarisch die Ergebnisse der Probanden aufgelistet. Die Tabelle zeigt neben dem Erfolg bei der Bearbeitung der Aufgaben und der erfassten Zuversicht auch die von den Probanden wahrgenommene Eignung der Darstellung der Ergebnisse. Die Frage zielte darauf ab, ob sich die Probanden von der Darstellung des Lernstandes repräsentiert fühlten (0 - „überhaupt nicht“, 7 - „vollkommen“). Die fünfte Spalte zeigt, welchen Fragenkatalog die Probanden für die Wiederholung verwendet haben. Die empfundene Nützlichkeit der Fragenkataloge wurde ebenfalls bewertet (0 – „ungeeignet“ bis 7 – „absolut geeignet“).

| Proband | Antwort-erfolg | Zuversicht | Repräsentation | verwendeter Fragenkatalog | Nützlichkeit |
|---|---|---|---|---|---|
| 1 | 45% | 55% | 6 | 1 | 7 |
| 2 | 69% | 65% | 6 | 4 | 7 |
| 3 | 57% | 53% | 5 | 1 | 6 |
| 4 | 50% | 69% | 6 | 1 | 7 |
| 5 | 65% | 74% | 6 | 4 | 6 |
| 6 | 77% | 84% | 6 | 4; 5 | k. A. |
| 7 | 64% | 51% | 3 | 3; 5; 6 | 6 |
| 8 | 84% | 54% | 6 | 5 | 7 |
| 9 | 74% | 62% | 5 | 4 | 5 |

Tab. 2: Ergebnisse der Evaluation des Fragenkataloges zur Prüfungsvorbereitung

Die Probanden wählten für die Prüfungsvorbereitung zwischen den drei Strategien: 1) alle Lernfragen noch einmal wiederholen, 4) alle nicht korrekt beantworteten Fragen, oder 5) alle Fragen mit einer geringen Zuversicht noch einmal bearbeiten. Das Minimieren von Wiederholungen (Katalog 2) und die Auswahl noch nicht bearbeiteter Fragen (Katalog 7) wurden von keiner Person gewählt. In den drei häufigsten Strategien findet sich Fragenkatalog 5, welcher als Grundlage die Zuversichtsbewertung für die Auswahl von Wiederholungsfragen nutzt. Aufgrund der geringen Probandenzahl können die Ergebnisse aber nicht als repräsentativ gewertet werde.

## 5 Zusammenfassung und Ausblick

Durch die Zuversichtserfassung ist es wesentlich leichter möglich, den Wissensstand der Teilnehmenden beim Beantworten von Lernfragen einzuschätzen. Dies kann in verschiedenen Kontexten wie dem Selbstregulierten Lernen (SRL) hilfreich sein, um die Lernstrategie anzupassen. Die zusätzliche Angabe der Zuversicht hinsichtlich der korrekten Bearbeitung der Aufgabe fördert die Reflexion des eigenen Wissensstandes und regt zu einer tieferen Auseinandersetzung mit der Lernfrage an. Darüber hinaus kann mit der Zuversichtsbewertung im Anschluss aufgezeigt werden, ob die Beurteilung des eige-

nen Wissenstandes korrekt war. Inwieweit die Zuversichtserfassung eine messbare Verbesserung des Lernerfolges oder eine Anpassung der Lernstrategie der Studierenden zur Folge hat, soll in weiteren Experimenten mit größeren Stichproben untersucht werden.

Literaturverzeichnis

[Br18] Braun, I. et al: Technology-enhanced self-regulated learning: Assessment support through an evaluation centre. Proc. IEEE 42nd Annual COMPSAC, Symposia CELT: Computer Education & Learning Technologies, Tokyo. 2018.

[Bu02] Burton, R. F.: Misinformation, partial knowledge and guessing in true/false tests. Medical Education, 36(9):805–811. 2002.

[Cu13] Curtis, D. A. et al: Does student confidence on multiple-choice question assessments provide useful information? Medical Education, 47(6):578–584. 2013.

[Ga06] Gardner-Medwin, A.: Confidence-Based Marking - towards deeper learning and better exams, Seiten 141– 149. Routledge. 2006.

[Hu03] Hunt, D. P.: The concept of knowledge and how to measure it. Journal of Intellectual Capital, 4(1):100– 113. 2003.

[Ka14] Kapp, F. et al: Meta-cognitive Support in University Lectures Provided via Mobile Devices - How to Help Students to Regulate Their Learning Process during a 90-minute Class. Proc. 6th CSEdu, Seiten 194–199. ScitePress. 2014.

[Ka11] Kapp, F.; Körndle, H.: Was lerne ich aus einer Lernaufgabe? a) gar nichts, b) Faktenwissen, c) etwas über meine Lernstrategien, d) Antwort b und c sind richtig. In T. Köhler & J. Neumann (Hrsg.), *Wissensgemeinschaften. Digitale Medien - Öffnung und Offenheit in Forschung und Lehre* (S. 178-187). Münster: Waxmann.

[Ku19] Kubica, T. et al: Choosing the appropriate audience response system in different use-cases. Proc. 10th ICETI, Orlando. 2019.

[LB16] Luetsch, K.; Burrows; J.: Certainty rating in pre- and post-tests of study modules in an online clinical pharmacy course - A pilot study to evaluate teaching and learning. In: BMC Medical Education 16.1, Seite 267. 2016.

[TT14] Tak, S.; Toet, A.: Color and Uncertainty: It is not always Black and White. In Elmqvist, N., Hlawitschka, M., Kennedy, J., editors, EuroVis - Short Papers. The Eurographics Association. 2014.

[WG10] Wakabayashi, T.; Guskin, K.: The Effect of an “Unsure” Option on Early Childhood Professionals’ Pre- and Post-Training Knowledge Assessments. In: American Journal of Evaluation 31.4, Seiten 486–498. 2010.

[Zi00] Zimmerman, B. J. et al: Attaining self-regulation: a social cognitive perspective. Handbook of self-regulation, 13. 2000.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 85

# Digitalisierungsprozesse an Hochschulen – der Blick der Akteur-Netzwerk-Theorie

Jana Niemeyer[1], Alexander Tillmann[2] und Michael Eichhorn[3]

**Abstract:** Digitalisierung ist, nicht nur im Bildungsbereich, in aller Munde: sie wird einerseits sehr positiv in Bezug auf neue Lehr-Lernmöglichkeiten und Teilhabe-Chancen diskutiert, andererseits auch stark in punkto Überwachung und Kontrolle kritisiert. Leugnen kann sie jedoch niemand. An den Hochschulen sind bereits viele Akteure mit der Digitalisierung der Hochschul(lehr)e befasst. Weitere werden durch Fördersummen in den kommenden Jahren hinzukommen. Wie sich die Digitalisierung dann tatsächlich ausgestaltet, wird ein spannender Prozess. Öffentliche Diskurse prägen dabei die Diskussionen mit, die an Hochschulen stattfinden. So sind die warnenden Stimmen genauso laut, wie diejenigen, die kritiklos(er) an die Thematik herangehen. Doch weder sozial- noch technikdeterministische Argumentationen werden langfristig die Diskussionen und Prozesse voranbringen. Die Betrachtung aller Akteure und ihre Verwicklungen in Netzwerken ist ein vielversprechender Schlüssel für eine aktive Mitgestaltung.

**Keywords:** Digitalisierung, Akteur-Netzwerk-Theorie, menschliche und nicht-menschliche Akteure, Teilhabe

## 1 Die Welle der Digitalisierung

Zwei Lager prägen aktuell den öffentlichen Diskurs über Digitalisierung und Bildung. Die einen, die Digitalisierung als unaufhaltsames Rad im Getriebe sehen und so schnell wie möglich Fortschritte im Bildungskontext sehen möchten, andere, die vor allem vor Gefahren der Ablenkung, Störung und Suchtverhalten warnen und Digitalisierungsprozessen nicht über den Weg trauen. Von Metaphern wie die Digitalisierungswelle überrollt uns aus der Ecke der Kritiker*innen oder der Zug darf nicht verpasst werden von Seiten der Befürworter*innen bis hin zu Wahlplakaten *Digitalisierung first, Bedenken second* auf bundespolitischer Ebene, ist alles mit dabei. Dass Digitalisierungsprozesse irgendwie einen Einfluss auf unser Leben und damit auch auf das Heranwachsen, Lernen und Lehren haben, wird dabei von keinem der Akteure geleugnet. Neuland ist es schon lange nicht mehr. Neben politischen Initiativen und Geldspritzen gibt es bereits viele Akteure, die seit Jahren an diversen Universitäten und Hochschulen Digitalisierungsprozesse gestalten, mal in kleinerem, mal in etwas größerem Kontext. Das Thema Digitalisierung schwappt (mal wieder) an die Hochschulen. „Jedoch scheint sich die Digitalisierungswelle am Bildungssystem zu brechen. Sie wird nicht mit offenen Armen aufgenommen oder gar umgesetzt“ [Br18, 179]. Häufig kommt in dem Zuge das Argument, die US-amerikanischen Hochschulen seien schon viel weiter: von dort könne

[1] Goethe-Univ. Frankfurt, studiumdigitale, Varrentrappstr. 40-42, 60486 FFM, niemeyer@sd.uni-frankfurt.de
[2] Goethe-Univ. Frankfurt, studiumdigitale, Varrentrappstr. 40-42, 60486 FFM, tillmann@sd.uni-frankfurt.de
[3] Goethe-Univ. Frankfurt, studiumdigitale, Varrentrappstr. 40-42, 60486 FFM, eichhorn@sd.uni-frankfurt.de

doi:10.18420/delfi2019_316

man viel lernen. Im selben Atemzug wird dann jedoch auf deren Sorglosigkeit im Umgang mit Daten verwiesen und auf unseren strengen Datenschutz (nicht erst seit EU-DSGVO), der viele Dinge einfach nicht möglich mache. Sehnsüchtig wird von manch Forscher*in in Digitalisierungsparadiese geschaut, die keinen strengen Datenschutz haben, zugleich aber kleinlaut beigeben, dass manch ethische Frage dort auch nicht gestellt werde. Digitalen Medien werden sogleich per se hohe Potentiale angedichtet [vgl. Br18, 181], kritische Stimmen fragen, vermutlich auch, weil sie die unkritische Behandlung neuer Tools ohne didaktische Einbettung kritikwürdig ist, nach tatsächlichen Beweisen der Wirkmächtigkeit gewisser Tools. Der Beweis eines Mehrwertes wird eingefordert [vgl. BH13] und betont, dass es keinerlei Hinweis darauf gäbe, dass der Einsatz digitaler Medien Lernleistungen erhöht. Gabi Reinmann plädiert dafür, dass die Hochschuldidaktik ihrerseits die Digitalisierung herausfordern sollte. “Die Fragerichtung ist nicht unerheblich, denn sie impliziert, wer das Sagen hat: Technik oder Bildung.” [Re19, 232] Sie möchte in dem Zuge auf die Verantwortung der hochschuldidaktischen Weiterbildung verweisen, die sich nicht dem Anschein eines digitalisierten Sachzwangs hingeben sollte [vgl. ebd.]. Die hier exemplarisch aufgeführten Dualismen können jedoch keine Richtschnur bei der Betrachtung digitaler Transformationsprozesse sein. Eine Antwort – so unsere Position – liegt in der Betrachtung der Akteur-Netzwerke. Diese Herangehensweise ermöglicht es, Dualismen aufzuheben und auf die emotionalen, teilweise polemisch geführten Pro-und-Contra-Diskussionen zu reagieren.

## 2 Die Akteur-Netzwerk-Theorie im Rahmen von Digitalisierungsprozessen

Während lange Zeit von technikdeterministischen oder sozialdeterministischen Ansätzen ausgegangen wurde, kamen in den 80er Jahren – gerade im Bereich der ‚Science and Technology’-Forschung – Ideen einer sozial konstruierten und endogenisierten Technikgeneseforschung auf. Das heißt weder der technologische Determinismus, noch der sozialdeterministische Ansatz seien die Antwort auf die Entwicklung, Steuerung und das Fortbestehen von Technologie in einer Gesellschaft, sondern die Kohäsion dieser beiden Ansätze [vgl. DW06, Ra99]. Der Technikdeterminismus beinhaltet die Vorstellung, dass technologische Änderungen und Entwicklungen unvermeidlich sind und kurzum Technologie nach dieser Auffassung Gesellschaft determiniert. Der Sozialdeterminismus nimmt hingegen Technologie resultierend aus der Gesellschaft wahr, in der sie entwickelt wird. Der Mensch entscheidet, wie sie sich entwickelt und wo sie zum Einsatz kommt [vgl. Be13, 3]. Seit den 80er Jahren kam es zu einer Verabschiedung von diesen dualistischen Ansätzen [Ca01, 5] und zu einem neuen Verständnis vom Zusammenspiel von Gesellschaft und Technologie, als ergebnisoffener und von sozialen Konstruktions-, Definitions- und Aushandlungsprozessen geprägter Prozess [vgl. DW06, 15]. Eine der bekanntesten Richtungen ist dabei die Akteur-Netzwerk-Theorie (ANT) [vgl. u.a. Ca87, La10].

Technik ist hierbei mitwirkender Agent (Aktant) in einem Netzwerk aus Beziehungen, das meint, dass Menschen wie Dinge als Mittler, Akteure in ihrem jeweiligen Netzwerk fungieren und, dass Technologien immer im Kontext des jeweiligen Netzwerkes zu sehen und zu beschreiben sind. Die ANT liefert dabei nicht nur theoretische Grundlagen, sondern auch methodische Orientierungen, da die Phänomene und Kontroversen der einzelnen Akteure und Aktanten innerhalb des Netzwerks gezielt beobachtet werden können [vgl. Pä16, 5]. „Netzwerke entstehen mittels Interaktionen, Transaktionen, Aushandlungen und Vermittlungen zwischen menschlichen und nicht-menschlichen Akteuren, die im Laufe dieser Prozesse bestimmte Rollen und Funktionen annehmen und ausführen. [...] Kommunikation zwischen Akteuren kann demzufolge als das Aushandeln eines >Interessenausgleichs< betrachtet werden." [BK06, 39] Jeder Akteur wird dabei nach dem "Symmetrieprinzip" betrachtet [vgl. La10, 184] und ist in seinem jeweiligen Netzwerk derjenige, der „eine gegebene Situation verändert, indem [er] einen Unterschied macht" [vgl. ebd., 123]. Die Situation würde folglich auch anders verlaufen, wenn ein „Akteur und seine Aktionen fehlen würde" [Pä16, 3]. Was bedeutet das nun für die Ausgestaltung und Initiierung von Digitalisierungsprozessen an Hochschulen?

## 3 Die ANT am Beispiel der Entwicklung des *Leitbilds Digitale Lehre* der Goethe-Universität Frankfurt

Hochschulen sind komplexe Organisationen, die sich im sozialwissenschaftlichen Verständnis als instrumentelles System von Regeln oder als institutionelles System handelnder Individuen beschreiben lassen. Die Alltagserfahrung ist jedoch viel stärker von materiellen Dingen wie Hörsälen, Laborgeräten, Beamern etc. geprägt. Die ANT versteht Universitäten daher als Akteur-Netzwerke, in denen menschliche und nicht-menschliche Akteure handeln [Pä16]. Die Frage, wie digitale Transformationsprozesse genau ausgestaltet werden sollen, wird an den deutschen Hochschulen kontrovers diskutiert. Diese inner-organisationalen Aushandlungsprozesse manifestieren sich oftmals in Strategiepapieren und Leitbildern. Auch die Goethe-Universität erarbeitete in einem breit angelegten Diskussionsprozess ein *Leitbild Digitale Lehre*, welches 2018 verabschiedet wurde [Go18]. In dem Papier lassen sich die oben beschriebenen Positionen zur Digitalen Transformation sowie die daraus resultierenden Konfliktlinien gut wiederfinden. Gleich der erste Satz zeigt allen Digitalisierungsbestrebungen klare Grenzen auf und gibt dem weiteren Text scheinbar die Linie vor: "Die [Goethe-Universität] bekennt sich als Präsenzuniversität zu einem zeitgemäßen Lehr- und Bildungskonzept" [Go18, 1]. Wenige Absätze später wird ein Anforderungsprofil an die Lehrenden formuliert, bei dem zumindest in Frage zu stellen ist, wie es sich im Rahmen einer "Präsenzuniversität" angemessen ausfüllen lässt: "Lehrende unterstützen Studierende [...] gleichermaßen beim Erwerb fachlicher wie digitaler Kompetenzen und haben im Umgang mit digitalen Medien eine Vorbildfunktion für ihre Studierenden." [Ebd., 2]

Die ANT bietet hier einen Erklärungsansatz für diese sich scheinbar widersprechenden Aussagen: Zum einen lassen sich eine Vielzahl menschlicher sowie organisationaler

Akteure identifizieren: die Lehrenden und Studierenden der Universität, ihre Mitarbeiter*innen in der Verwaltung sowie die durch die handelnden Personen repräsentierten organisationalen Einheiten wie Institute oder zentrale Einrichtungen. Daneben lassen sich in dem an der Erstellung des Leitbilds beteiligten Akteurs-Netzwerk "Organisation Universität" aber auch nichtmenschliche Akteure mit teils gegenläufigen Interessen finden. Als ein solcher Akteur könnten z.B. die erst wenige Jahre alten Campusgebäude mit zahlreichen neuen Hörsälen, Seminar- und Laborräumen betrachtet werden, welche eine möglichst intensive Nutzung im Sinne einer "Präsenzuniversität" sinnvoll erscheinen lassen. Auf der anderen Seite stellen bereits vorhandene digitale Infrastrukturen wie bspw. Videosysteme zur Erstellung von Vorlesungsaufzeichnungen oder Lernmanagement-Systeme wirkmächtige Akteure dar, die Digitalisierungsprozesse befördern. Nicht zuletzt wird das *Leitbild* selbst Teil dieses Akteur-Netzwerks.

## 4 Die ANT als Hilfsmittel für den Digitalisierungsstrategie-Prozess

Gerade durch die Betrachtung aller relevanten (nicht-menschlichen) Akteure kann die ANT für Universitäten eine wertvolle Planungshilfe bei der Erarbeitung einer Digitalisierungsstrategie sein. Hierbei lassen sich sowohl materielle wie immaterielle Akteure identifizieren. Der Mehrwert besteht dabei vor allem darin, den in den Verbindungen und Netzwerkbeziehungen integrierten Dingen (digitale Medien, ICT, Gebäude etc.) *Handlungsmacht* und *Handlungspotential* zuzugestehen. Das Vokabular des ANT-Ansatzes dient dabei als Instrumentarium um die Relationen der Akteure offenzulegen, Hindernisse und Fördermechanismen im Rahmen von Digitalisierungsprozessen und den Umgang mit digitalen Medien in Lehr-/Lernprozessen zu analysieren, ohne diese vorschnell als pädagogisch relevant oder irrelevant ein- oder auszusortieren.

Um Prozesse des Scheiterns oder Gelingens der Integration digitaler Medien und ICT in der Hochschule oder Schule zu beforschen, stellt die ANT eine Herangehensweise dar, die sozialen Wirklichkeiten und Entwicklungen innerhalb eines Akteur-Netzwerkes zu analysieren und zu beschreiben. Wird die Art der Leistungsbeurteilung beispielsweise ebenfalls als ein (immaterieller) Akteur betrachtet, so wird deutlich, dass diese z.B. einen erheblichen Einfluss auf die Art und Weise der Nutzung digitaler Medien ausübt. Bei der Beobachtung der sozialen Wirklichkeiten der Integration digitaler Medien in einem konkreten Lehr-/Lernkontext kann dann z.B. offengelegt werden, dass das der Leistungsbeurteilung eingeschriebene Handlungsskript völlig different von den Handlungsskripten aller anderen Akteure im Netzwerk ist. Die Beobachtung dieser Inkongruenzen kann dann empirisch aufzeigen, warum das spezifische Akteur-Netzwerk eine anvisierte Wirkung (z.B. einen über die Leistungsbeurteilung gemessenen Lerneffekt) wahrscheinlich nicht abbilden wird. In ähnlicher Weise fruchtbar kann die ANT bei der Identifizierung von Bildungsbarrieren gemacht werden. Die empirische Untersuchung von differenten und kongruenten Handlungsskripten von Akteuren in einem konkreten Akteurs-Netzwerk zeigt z.B. Potenziale digitaler Medien bezogen auf In- und

Exklusion bestimmter Zielgruppen auf. Dazu müsste zunächst möglichst breit erhoben werden, welche Akteure in dem Netzwerk und welche Erwartungen und Beziehungen der Akteure auszumachen sind. Um Bildungsbarrieren zu identifizieren kann im Anschluss gefragt werden, welche differenten Erwartungszuschreibungen beobachtet werden können. Auf dieser Grundlage können dann Implementierungsstrategien digitaler Medien entworfen werden, mit der Intention Bildungsbarrieren abzubauen und Teilhabe zu ermöglichen. Akteure mit Rollenzuschreibungen, Beziehungen und Erwartungen, die dabei in den Blick genommen werden sollten sind z.B. Studierende, Lehrende, das Curriculum, Lernplattformen, Autorensysteme, Infrastruktur, Förderprogramme, didaktische Fort- und Weiterbildung und Datenschutzverordnungen. Wie in dieser Gemengelage einer Bildungseinrichtung als Netzwerk von Akteuren die Rolle der Technik in Bildung konzeptualisiert wird, ist entscheidend, da je nach Verständnis dieser Rolle unterschiedliche Handlungsprogramme und Strategien auf Seiten der sozialen Akteure resultieren. Ziele werden anders gesetzt, menschliche, technische und finanzielle Ressourcen zugesprochen oder nicht, künftige Entwicklungen durch strategische Entscheidungen initiiert, Rahmenbedingungen für gesellschaftliche Änderungen gesetzt und entsprechende Forderungen an alle Beteiligten im Bildungssystem gestellt [vgl. Be13, 2]. Mithilfe der ANT können diese Prozesse beobachtet und Abhängigkeiten im spezifischen Kontext besser verstanden werden.

## 5 Fazit

Digitalisierung kann sehr gut in Akteur-Netzwerken gedacht, beobachtet und beforscht werden. Die Frage ist, ob wir die Prozesse als hybrides Konstrukt begreifen, um Digitalisierung besser einzuordnen, zu beschreiben und zu gestalten, sie weder als Gefahr noch als Heilsbringer zu deklarieren. Es sind die Aushandlungsprozesse, das Dazwischen, der gezielte Einsatz didaktischer Fragestellungen in kombinatorischer Betrachtung diverser Akteure und Aktanten, mit denen wir (als medienpädagogische Instanz) interagieren und sie als solche anerkennen. „Die Akteur-Netzwerk-Theorie kann helfen, einen offeneren Zugang zur Beschreibung von Lernkulturen zu erlangen, der auch Medien adäquat verortet.“ [vgl. La13, 180]. Unsere Position dazu ist, die aufgeheizte Diskussion abzukühlen und mit einem nüchternen wissenschaftlichen Blick, die Digitalisierungsprozesse medienpädagogisch zu begleiten, zu hinterfragen, zu verfolgen und mitzugestalten. Es lohnt sich dabei die ANT zu beachten und für empirische und organisationale Prozesse in den Blick zu nehmen.

Literaturverzeichnis

[Be13] Belliger, A. et.al.: Die Akteur-Netzwerk-Theorie - Eine Techniktheorie für das Lernen und Lehren mit Technologien. In (Ebner, M.; Schön, S. Hrsg.): Lehrbuch für Lernen und Lehren mit Technologien. 2. Auflage, ePubli GmbH, Berlin, S. 335-342, 2013.

[BH13] Baumgartner, P. & Herber, E.: Höhere Lernqualität durch interaktive Medien? - Eine kritische Reflexion. 2013

[BK06] Belliger, A.; Krieger, D. J. (Hrsg.): ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Transcript Verlag, Bielefeld, 2006.

[Br18] Braungardt, K.: Digitalisierungsprozesse zwischen Anpassung und Kritik. Überlegungen zu digitalisierungskritischen Praktiken in der Hochschule. In (Leineweber, C.; de Witt, C., Hrsg.): Digitale Transformation im Diskurs. Kritische Perspektiven auf Entwicklungen und Tendenzen im Zeitalter des Digitalen. deposit_hagen Publikationsserver der Universitätsbibliothek, Hagen, S. 179-196, 2018.

[Ca87] Callon, M.: Society in the Making: The Study of Technology as a Tool for Sociological Analysis. In (Bijker, W. E.; Hughes, T. P.; Pinch, T. J. Hrsg.): The Social Construction of Technological Systems. MIT Press Cambridge MA, London, S. 83-103, 1987.

[Ca01] Castells, M.: Der Aufstieg der Netzwerkgesellschaft. VS Verlag für Sozialwissenschaften, Wiesbaden, 2001.

[DW06] Dolata, U.; Werle, R.: Bringing technology back in: Technik als Einflussfaktor sozioökonomischen und institutionellen Wandels. Campus-Verlag, Frankfurt, 2006.

[El17] Elm M.: Technology teaches Action: Ästhetische Erfahrungen hybrider Akteure Überlegungen zur Formierung neuer Orientierungen Erwachsener in Auseinandersetzung mit der Stimmung neuer Medien. In: Münk D., Walter M. (eds) Lebenslanges Lernen im sozialstrukturellen Wandel. Springer VS, Wiesbaden, S. 171-183, 2017.

[Go18] Goethe-Universität: Leitbild digitale Lehre an der Goethe-Universität, 2018, http://www.uni-frankfurt.de/72312239/RZ_Leitbild_dig_Lehre_A4_low.pdf, Stand: 13.03.2019

[La10] Latour, B.: Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie. Suhrkamp, Berlin, 2010.

[La13] Laschewsky, J.: Vernetzte Lernkultur – Eine Perspektivenerweiterung durch die Akteur-Netzwerk-Theorie nach Bruno Latour. In (Dollhausen, K.; Feld, T.; Seitter, W. Hrsg.) Erwachsenenpädagogische Kooperations- und Netzwerkforschung. Theorie und Empirie Lebenslangen Lernens. Springer VS, Wiesbaden, S. 171-184, 2013.

[Pä16] Pätzold, H.: Akteur-Netzwerk-theoretische Grundlagen der Organisationspädagogik. In (Göhlich, M.; Schröer, A.; Weber, S. M. Hrsg.): Handbuch Organisationspädagogik. Springer Fachmedien, Wiesbaden, S. 1–11, 2016.

[Ra99] Rammert, W.: Technik. Stichwort für eine Enzyklopädie. Technical University Technology Studies. TU-Berlin, Berlin, 1999.

[Re19] Reinmann, G.: Digitalisierung und hochschuldidaktische Weiterbildung: Eine Kritik. In (Heider-Lang, J.; Merkert, A. Hrsg.): Digitale Transformation in der Bildungslandschaft den analogen Stecker ziehen? Rainer Hampp Verlag, Augsburg, München, 2019.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 91

# Technische Aspekte der automatischen Aufgabengenerierung für *Blended Learning* Angebote in der Biologie

**Eine Analyse an zwei Fallbeispielen aus der Botanik und Genetik**

Justin Timm[1], Benjamin Otto[2], Thilo Schramm[1], Michael Striewe[2], Philipp Schmiemann[1] und Michael Goedicke[2]

**Abstract:** Der Beitrag demonstriert und analysiert an zwei Fallbeispielen aus der Biologie, wie fachspezifische Selbstlernaufgaben mit variablen Inhalten für ein *Blended Learning* Angebot automatisiert erzeugt werden können. Es wird gezeigt, dass auch hochspezifische fachliche Eigenschaften technisch abgebildet werden können und eine gute Item-Qualität erzeugt werden kann.

**Keywords:** Blended Learning, E-Assessment, Aufgabengenerierung, Biologiedidaktik

## 1 Einleitung

*Blended Learning* beschreibt die Integration von E-Learning-Angeboten in bestehende Präsenzveranstaltungsformate. Durch die Kombination der Vorzüge beider Formate kann der Kompetenzerwerb in den entsprechenden Veranstaltungen gesteigert werden [BB14, Me13], wobei die Entwicklung der Inhalte für die einzelnen Lernumgebungen maßgeblich von der Struktur der Fachdisziplinen bestimmt wird. Die Biologie als Fachwissenschaft zeichnet sich durch eine Vielzahl an Zweigen und Disziplinen aus, die in der Regel ihre eigenen weit zurückreichenden Forschungstraditionen aufweisen [Ma00], in denen häufig auch auf disziplinenspezifische Lehrmethoden zurückgegriffen wird. Im Rahmen dieses Beitrags sollen technische Aspekte bei der automatisierten Aufgabengenerierung für Lehrveranstaltungen aus zwei unterschiedlichen Fachdisziplinen der Biologie vorgestellt werden. Dies ist zum einen die *botanische Übung zur Biodiversität*, die der allgemeinen Botanik zuzuordnen ist und im Weiteren als *botanische Bestimmungsübung* bezeichnet wird, und die *Einführung in die Genetik*.

In beiden Fällen werden fachspezifische Leistungsanforderungen an die Studierenden gestellt, von denen viele ausschließlich in disziplinenspezifischen Aufgaben abgebildet werden können. In der *botanischen Bestimmungsübung* sollen Studierende beispielsweise

[1] Universität Duisburg-Essen, Fakultät für Biologie, Didaktik der Biologie, Universitätsstraße 2, 45141 Essen, {justin.timm,thilo.schramm,philipp.schmiemann}@uni-due.de
[2] Universität Duisburg-Essen, paluno - The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, {benjamin.otto,michael.striewe,michael.goedicke}@paluno.uni-due.de

doi:10.18420/delfi2019_262

lernen, Blütenpflanzen anhand morphologischer Merkmale zu bestimmen, wozu im Rahmen solcher Übungen in der Regel der Umgang mit einem Bestimmungsschlüssel eingeübt wird. In der Genetik wird von den Studierenden unter anderem erwartet, dass sie einen Vererbungsmechanismus eines genetisch determinierten Merkmals anhand des Auftretens innerhalb einer Familie aufklären. Hierzu existiert der fachspezifische Abbildungstypus des Familienstammbaums, der konventionalisiert ausschließlich zur Beantwortung genetischer Fragestellungen eingesetzt wird. In beiden Fällen ist dabei deklaratives Wissen (über Blütenpflanzen beziehungsweise Vererbungsmechanismen) und prozedurales Wissen (über den Umgang mit einem Bestimmungsschlüssel beziehungsweise das logische Schlussfolgern auf Grundlage von Vererbungsregeln und Wahrscheinlichkeiten) notwendig. Ziel der entwickelten *Blended Learning* Angebote ist es, die Entwicklung und Sicherung prozeduraler Fähigkeiten zu unterstützen.

Überprüft wird im Rahmen dieses Beitrags, inwiefern es gelingt, automatisch aus hochspezifischen Aufgaben(-Formaten) Übungsaufgaben zu generieren, die auch beim wiederholten Einsatz nicht durch bloßes Memorieren zu lösen sind. Zentrale Fragestellungen des Beitrags sind dabei, wie unter Berücksichtigung der fachspezifischen Besonderheiten eine hohe Item-Qualität sichergestellt werden kann und an welcher Stelle des Generierungsprozesses biologisches Fachwissen erforderlich ist. Exemplarisch vorgestellt wird hierzu eine Aufgabe zur Identifikation einer Blütenformel in Abhängigkeit von einer gegebenen Pflanzenfamilie für die *Bestimmungsübungen* und eine Aufgabe zur Analyse eines Familienstammbaums zur Bestimmung des vorliegenden Vererbungsmechanismus im Rahmen der *Einführung in die Genetik*.

## 2 Verwandte Arbeiten

Grundsätzlich ist die automatische Aufgabengenerierung ein gut beforschter Bereich mit zahlreichen existierenden Ansätzen [GLZ18]. Insbesondere im Kontext von Multiple-Choice-Aufgaben kann sie gezielt zur Steigerung der Qualität von Distraktoren verwendet werden [La16]. Fachspezifische Ansätze aus der Biologie sind jedoch wenig verbreitet. Hier kommen eher manuell erstellte Aufgaben in geschlossenen Lern-Umgebungen (z. B. [Co17]) zum Einsatz. In anderen Fächern (z.B. der Mathematik [Ku15, Sa13]) ist dagegen die Nutzung fachspezifischer Systeme auch für andere Aufgabentypen verbreitet.

Prinzipiell domänenunabhängig funktionieren dagegen Ontologie-Ansätze, zu denen es mannigfaltige Forschung gibt; ein detailliertes Literature Review findet man in [Al15]. Beispielhaft nutzt [SB14] Formal Concept Analysis (FCA), um Elektrotechnik-Aufgaben zu generieren und Alsubait et al. evaluieren verschiedene Ansätze [Al12] zur Ontologie-Generierung von Multiple-Choice-Fragen, die auch Biologie-Ontologien umfassen. Diese Ansätze sind allerdings kaum für die konkreten Fallbeispiele in diesem Beitrag geeignet, da sich die Kombination aus prozeduralem Wissen in Kombination mit falschem deklarativem Wissen als Distraktor nicht unmittelbar aus Ontologien ableiten lässt.

## 3 Technischer und organisatorischer Hintergrund

Das im Rahmen dieses Beitrags verwendete E-Assessment-System JACK ermöglicht parametrisierte Aufgaben, bei denen sich Inhalte bei jedem Aufruf der Aufgabe ändern. In der Mathematik sind solche Verfahren üblich, um dieselbe Aufgabe mit unterschiedlichen Zahlenwerten anzubieten. Das Prinzip ist jedoch nicht auf Zahlenwerte beschränkt, sondern kann auch Texte, Bilder oder andere Aufgabenbestandteile betreffen. Die Parametrisierung fördert Studierende so, theoretische Konzepte zu durchdringen, denn einerseits ist Auswendiglernen bei Aufgaben mit genügender Variabilität nicht mehr möglich, andererseits können diese auch nach erfolgreichem Abschluss mehrfach mit anderen Werten geübt werden, um das Verständnis des Stoffes zu vertiefen.

Das E-Assessment-System JACK enthält eine Auswertungseinheit, mit der Variablen in Aufgaben sowohl symbolisch als auch numerisch durch verschiedene Ausdrücke und Funktionen, aber auch durch andere Verfahren wie Ontologieabfragen belegt werden können. Nach Bedarf kommen dazu intern Computer-Algebra-Systeme (CAS) wie Sage oder die statistische Programmiersprache R zum Einsatz.

Die *botanische Bestimmungsübung* wird in jedem Sommersemester angeboten und jeweils von ca. 250 Studierenden besucht. In der Veranstaltung werden an jedem Kurstag Pflanzenfamilien und ihre morphologischen Merkmale vorgestellt und mithilfe eines Bestimmungsbuchs anhand echter Pflanzen bestimmt. So soll zum einen deklaratives Wissen über Bau und Taxonomie der Pflanzen erworben werden, zum anderen aber auch dessen Anwendung im praktischen Bestimmen. Die Veranstaltung schließt mit einer Klausur ab, welche aus einem theoretischen Teil über Bestimmungsmerkmale und Pflanzenfamilien und einem praktischen Teil, in welchem Studierende vier gegebene Pflanzen korrekt bestimmen müssen, besteht. Zur Begleitung der Bestimmungsübung und zur Vorbereitung auf die Modulabschlussprüfung wurde im Sommersemester 2018 erstmals eine Online-Lernumgebung bereitgestellt. In insgesamt 175 Aufgaben konnten Studierende den theoretischen Teil der Übung in Themen wie beispielsweise Blütenformeln, Blütendiagrammen, Blütenaufbau, Pflanzenaufbau, Fruchtformen oder wissenschaftlicher Nomenklatur üben.

Im Wintersemester 2018/19 wurde die Vorlesung *Einführung in die Genetik*, die im Rahmen des Bachelor-Studiums von Lehramtsstudierenden sowie den Studierenden der Biologie und der medizinischen Biologie belegt werden muss ($N_{ges} \approx 300$), erstmals durch ein digitales Lernangebot unterstützt. Es sollte zur Nachbereitung der Vorlesung und Vorbereitung auf die Klausur von den Studierenden eingesetzt werden können und ein stattfindendes Tutorium ergänzen und entlasten. Das digitale Lernangebot umfasste insgesamt 32 dynamisch generierte Übungsaufgaben zur Formal-, Human- und Populationsgenetik, von denen 20 bereits empirisch erprobt wurden [SNT17]. In Fallbeispiel 2 wird eine neue Aufgabe zur Stammbaumanalyse vorgestellt.

# 4 Fallbeispiel 1: Botanik

Die automatische Generierung von Aufgaben stellt in der Biologie grundsätzlich eine Herausforderung dar, da in Aufgaben typischerweise Faktenwissen thematisiert wird. Es finden sich in den meisten keine Formeln, welche wie in der Mathematik einfach variiert werden können, um so einen Aufgabenpool zu generieren. Bei Beschriftungs- oder Zuordnungsaufgaben kann eine Variation durch die verwendeten Bilder und die erfragten Strukturen oder Prozesse erfolgen, jedoch sind auch diese Variationsmöglichkeiten sehr begrenzt. Eine Chance bietet sich jedoch im speziellen Fall der Botanik in der Bestimmungsübung, in der unter anderem sogenannte Blütenformeln eingesetzt werden.

Damit Studierende gegebene Pflanzen ihrer jeweiligen Familie zuordnen können, müssen sie verschiedene Merkmale der Familien kennen. Eines der zentralen Merkmale ist hierbei der Aufbau der Blüte. Die komprimierteste Darstellung des Blütenaufbaus ist die Blütenformel, das Wissen um die Blütenformel reicht in vielen Fällen aus, um die Familie einer Pflanze anhand der Blüte zu bestimmen. Blütenbestandteile werden von außen nach innen durch Buchstaben und Zahlen bezeichnet. Die Anzahl, Anordnung und gegebenenfalls Verwachsungen der Bestandteile werden durch Zahlen, Klammern und weitere Symbole dargestellt. So lautet die Formel der Familie *Caryophyllaceae* (Nelkengewächse): * K(5) C5 A5+5 G(5) und gibt an, dass es sich um eine radiärsymmetrische Blüte (*, also min. 3 Symmetrieachsen) mit 5 verwachsenen Kelchblättern (K(5)), 5 Kronblättern (C5), 5 äußeren und 5 inneren Staubblättern (A5+5) und 5 verwachsenen, oberständigen Fruchtblättern (G(5)) handelt.

Im Rahmen der Lernplattform wurde ein Aufgabenformat entwickelt, bei welchem Studierende die richtige Blütenformel zu einer gegebenen Pflanzenfamilie angeben müssen. Ihnen wird dazu nur der Name der Pflanzenfamilie vorgegeben, aber keine zugehörige Abbildung gezeigt.

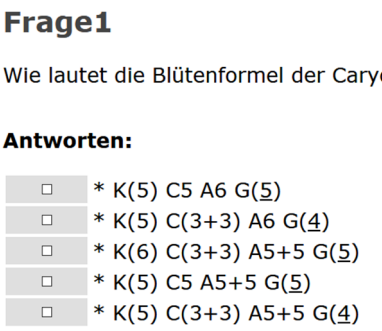



Abb. 1: Beispielaufgabe zu Blütenformeln (Live-Demo verfügbar unter https://jack-demo.s3.uni-due.de/jack2/demo?id=68204)

Für die automatische Erzeugung der Distraktoren wird ein zweistufiges Zufallsverfahren verwendet. Im ersten Schritt werden eine zufällige, falsche Zahl und zwei zufällige falsche

Summen ausgewürfelt, wobei von den Summen eine garantiert aus zwei gleichen Summanden besteht. „Falsch“ bedeutet in diesem Kontext, dass Zahlenwerte verwendet werden, die in der richtigen Lösung nicht auftreten, aber aus fachlicher Sicht nicht völlig abwegig erscheinen. An dieser Stelle des Konstruktionsprozesses ist also botanisches Fachwissen notwendig, um die Wahl völlig abwegiger Werte zu vermeiden. Auf Basis dieser Zufallswerte werden anschließend für alle vier Bestandteile der Blütenformel mehrere falsche Inhalte erzeugt, indem die falschen Zahlen, falschen Summen und ggf. noch zusätzliche falsche Klammerungen oder Über-/Unterstriche verwendet werden, soweit dies fachlich sinnvoll ist. An dieser Stelle des Konstruktionsprozesses ist daher ebenfalls botanisches Fachwissen notwendig, um die Konstruktion unsinniger Distraktoren ausschließen zu können. Im Beispiel aus Abbildung 1 entstehen so für die Bestandteile K und C jeweils 9 falsche Möglichkeiten, für A sechs Möglichkeiten und für G fünf Möglichkeiten. Aus dieser Menge von falschen Bestandteilen werden anschließend im zweiten Zufallsschritt die Distraktoren gebildet, wobei ein Distraktor aus genau einem falschen und drei richtigen Bestandteilen zusammengesetzt wird, zwei aus genau zwei falschen Bestandteilen und einer aus genau drei falschen Bestandteilen. Durch die Zweistufigkeit des Verfahrens ist dabei sichergestellt, dass die Distraktoren nicht schon alleine dadurch zu identifizieren sind, dass sie immer andere Zahlenwerte enthalten. Im Beispiel aus Abbildung 1 kommt etwa der falsche Bestandteil C(3+3) häufiger vor als der korrekte Bestandteil C5. Die Aufgabe lässt sich also nicht einfach dadurch lösen, dass man jene Option wählt, die den jeweils häufigsten Wert für jeden Bestandteil enthält.

Im obigen Beispiel ergeben sich rechnerisch insgesamt 320 mögliche Kombinationen für falsche Zahlenwerte und für jede dieser Kombinationen können 2.430 Distraktoren erzeugt werden. Insgesamt können bei dieser Aufgabe also 777.600 verschiedene Distraktoren erzeugt werden, wobei nicht jede beliebige Kombination aus dieser Menge innerhalb einer Aufgabe auftreten kann. Auch wenn die Distraktoren gleichverteilt gezogen werden, ist zudem das Auftreten einzelner Bestandteile nicht gleichverteilt, da die Mengen an potenziellen Bestandteilen unterschiedlich groß sind. Es könnte daher in ungünstigen Fällen theoretisch passieren, dass schlechtere Distraktoren häufiger auftreten. Dieser Frage soll im Folgenden durch die exemplarische Analyse von Nutzungsdaten aus dem Wintersemester 2017/18 nachgegangen werden.

Die oben gezeigte Beispielaufgabe zur Blütenformel der *Caryophyllaceae* wurde im Untersuchungszeitraum von 104 Studierenden insgesamt 279 Mal bearbeitet. Insgesamt wurden den Studierenden dabei 795 verschiedene Distraktoren angezeigt. Die korrekte Antwort wurde in 234 Fällen gewählt (84%), was dem mittleren Schwierigkeitsniveau aller eingesetzten Aufgaben dieses Typs entspricht. Im ersten Versuch wurde die Aufgabe in 79% der Fälle gelöst. Die drei am häufigsten angezeigten Distraktoren stammen alle aus der Gruppe der Distraktoren mit einem falschen Bestandteil und wurden insgesamt 46 Mal angezeigt und in 9 Fällen gewählt (20%). Ein weiterer exemplarischer Distraktor mit einem falschen Bestandteil wurde 9 Mal angezeigt und 4 Mal gewählt. Im Gegenzug gibt es jedoch auch einen Distraktor mit einem falschen Bestandteil, der 12 Mal angezeigt aber nie gewählt wurde. Distraktoren mit drei falschen Bestandteilen wurden insgesamt nur ein

einziges Mal gewählt. Die automatische Generierung führt demnach prinzipiell zu guten Ergebnissen, produziert allerdings auch viele überflüssige Distraktoren.

Das Ergebnis ist jedoch nicht auf andere Blütenformeln verallgemeinerbar. Eine andere Aufgabe zur Blütenformel der *Fabaceae* wurde von 90 Studierenden 203 Mal bearbeitet. Insgesamt wurden 609 Distraktoren angezeigt und die korrekte Antwort in 174 Fällen (86%) gewählt, so dass die Aufgabe prinzipiell vergleichbar ist. Die drei häufigsten Distraktoren wurden jedoch insgesamt 36 Mal angezeigt, aber nur einmal gewählt (3%). Hier liegt also tatsächlich eine Schwäche bei der Erzeugung der Distraktoren vor, obwohl das Konstruktionsprinzip der Aufgabe identisch ist.

Aus diesen Werten lassen sich mehrere Schlussfolgerungen mit Blick auf die Forschungsfragen dieses Papers ziehen: Das erste Beispiel zeigt, dass die automatische Erzeugung guter Distraktoren prinzipiell möglich ist. Beide Beispiele zusammen zeigen, dass eine große Zahl von Distraktoren erzeugt wird, die nicht oder nur selten gewählt wird und damit zu eliminieren sind. Es erscheint daher notwendig, den Prozess der Aufgabenerstellung und Qualitätssicherung iterativ zu gestalten, indem die Konfiguration der Erzeugung von Antwortoptionen schrittweise so angepasst wird, dass häufig auftretende aber selten gewählte Distraktoren für künftige Einsätze nicht mehr erzeugt werden. Anders als bei der initialen Festlegung der potenziellen falschen Bestandteile der Distraktoren ist für die iterative Verfeinerung jedoch nur ein geringes Maß an botanischem Fachwissen notwendig.

Im Rahmen der ersten Nutzung im Sommersemester 2018 haben 197 Studierende das Lerntool genutzt und dabei insgesamt über 24.000 Aufgaben gelöst. Im Laufe des Semesters wurden typischerweise an den Tagen vor den Testaten 600-900 Aufgaben pro Woche gelöst mit einem starken Anstieg vor der Klausur. Dies deutet darauf hin, dass die Studierenden die Plattform als lernförderlich erachten. Genauere Analysen mit Blick auf Auswirkungen der Nutzung auf Klausurergebnisse sind für das Sommersemester 2019 geplant.

## 5 Fallbeispiel 2: Genetik

Humangenetische Stammbäume stellen das Vorkommen spezifischer in der Regel erblicher Merkmale innerhalb familiärer Zusammenhänge dar. Sie können zur Bestimmung des dem Merkmal zugrundeliegenden Erbgangs, bei der Kartierung von Erbkrankheiten [Gr15, S. 604], aber auch im Rahmen einer humangenetischen Beratung eingesetzt werden, wobei das Ziel der genetischen Beratung sowohl therapeutische, als auch präventive Maßnahmen sein können [TG11, S. 46].

Die Analyse humangenetischer Stammbäume zur Identifikation des dargestellten Vererbungsmechanismus ist erstmalig bereits im Biologieunterricht der weiterführenden Schule Gegenstand des Unterrichts (vgl. [KMK05, KMK04]). Im Rahmen der universitären Lehrveranstaltung *Einführung in die Genetik* wird von den Studierenden

erwartet, dass sie sowohl auf Grundlage eines gegebenen Stammbaums den Vererbungsmechanismus eines Merkmals aufklären, als auch die Wahrscheinlichkeit ermitteln können, mit der ein im Stammbaum dargestelltes Elternpaar ein erkranktes Kind bekommt. Hierbei deuten bisherige Studienergebnisse an, dass die Schwierigkeit von "Stammbaumaufgaben" maßgeblich vom dargestellten Vererbungsmechanismus abhängig ist (vgl. [Ha94]).

Das JACK-R-Modul [Ot17] stellte eine geeignete Basis dar, um parametrisierte Genetikaufgaben mit dynamisch generierten, grafischen Stammbäumen stellen zu können, da mit *kinship2* [STS14] bereits ein R-Paket zur grafischen Darstellung von Stammbäumen (siehe Abb. 2) im *Comprehensive R Archive Network (CRAN)* zur Verfügung stand. Zur Nutzung derartiger Pakete gibt es in JACK verschiedenartige Anbindungen an sogenannte Backends, die R-Code auswerten können. Insbesondere war bereits ein auf R-Serve [Ur03] basierender Service vorhanden, der Auswertungen von R-Programmen zurückgibt. Um dabei die Ausführung von potentiell unsicherem R-Code sicher zu gestalten, wird dieser isoliert in Docker-Containern ausgeführt, welche zusätzlich noch mit Blacklisting [Da13] von sicherheitsrelevanten R-Befehlen arbeiten. Um auch die für Stammbaumaufgaben notwendigen Bilder einbinden zu können, wurde zusätzlich ein auf Spring basierender Microservice entwickelt, der versucht, aus übergebenen R-Befehlen Bilder zu erstellen, und diese im Erfolgsfall JACK bereitstellt. Zusätzlich stellt ein neu entwickeltes R-Paket die notwendigen Funktionen zur Konstruktion der Aufgaben bereit, die im Folgenden näher erläutert werden.

Da es sich bei Stammbaumaufgaben um Problemlöseaufgaben handelt, erfordert deren korrekter Lösung in der Regel die Integration von fachlichen und fachübergreifenden Fähigkeiten (vgl. [TT10, S. 1078]). Um den Studierenden ausreichend Übungsmöglichkeit zu schaffen, wurden automatisiert Stammbäume zu allen lehrveranstaltungsrelevanten Vererbungsmechanismen erzeugt und so halbautomatisiert eine große Zahl von Stammbaumaufgaben generiert. Die Aufgabenkonstruktion läuft in drei Schritten ab, wobei die Generierung des Stammbaums die ersten zwei Schritte in Anspruch nimmt.

In einem ersten Schritt wird nach bestimmten Regeln eine Familienkonstellation erstellt. Dabei wird ausgehend von einer Elterngeneration ein mehrstufiges und in Teilen iteratives Zufallsverfahren eingesetzt, um eine realistische Familienstruktur zu generieren. In den erzeugten Stammbäumen werden dabei stets zwischen drei und fünf Generationen dargestellt, die auf ein Gründerpaar zurückgehen, welches mindestens drei und maximal fünf Kinder besitzt. Ab der zweiten Generation werden dann zufallsbasiert jeweils zwischen zwei und vier Personen ausgewählt, für die ein nicht verwandter Partner beziehungsweise eine nicht verwandte Partnerin und zwei ($P = 3/6$), drei ($P = 2/6$) oder vier ($P = 1/6$) Kinder erzeugt werden.

Im zweiten Schritt wird auf Grundlage des zu simulierenden Erbmechanismus zufallsbasiert die Vererbung der merkmalsbestimmenden Genkopien (Allele) vollzogen. Hierzu muss zunächst die genetische Ausstattung des Gründerpaars und der in jeder Generation neu hinzugefügten Partner in Abhängigkeit vom simulierten Erbgang

zufallsbasiert generiert werden. Erst im Anschluss können die Allele den Regeln des jeweiligen Erbmechanismus folgend, an die nachkommenden Generationen weitergegeben werden. Wobei auch hier in der Regel eine Zufallsfunktion zum Tragen kommt, da bei einer nicht geschlechtsgebundenen Vererbung jeder Mensch von jedem Elternteil genau eine von zwei Genkopien erbt. Danach lässt sich auf Grundlage der vorliegenden beziehungsweise ererbten Allele und des zu simulierenden Erbmechanismus bestimmen, welche Person das Merkmal trägt (und damit in der grafischen Darstellung durch ein eingefärbtes Symbol repräsentiert wird), womit die Generierung des Familienstammbaums abgeschlossen ist.

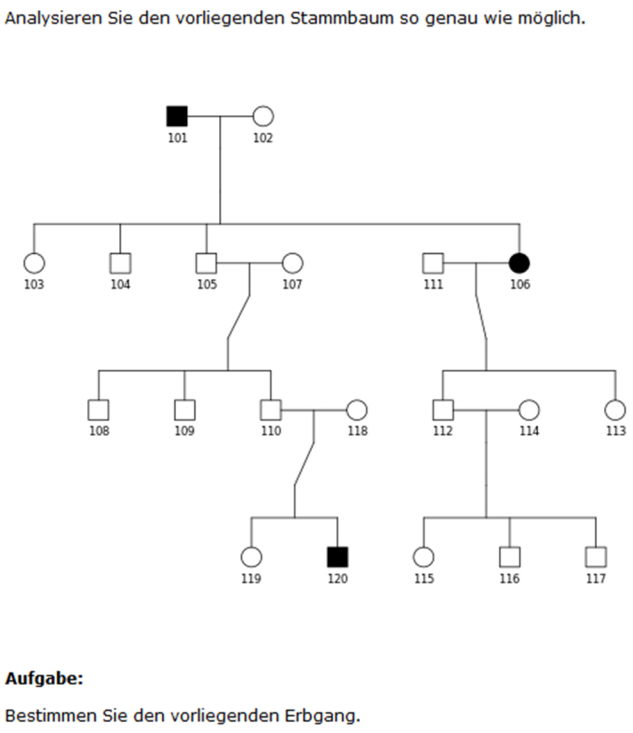



Abb. 2: Stammbaumaufgabe auf Grundlage eines autosomal-rezessiven Erbgangs. Innerhalb des Stammbaums werden männliche Personen durch Quadrate und weibliche Personen durch Kreise dargestellt. Linien visualisieren die familiäre Beziehung zwischen den dargestellten Personen. Ausgefüllte Symbole repräsentieren Betroffenheit einer Person mit dem jeweiligen Merkmal. (Nomenklatur bei [Be08])

Obwohl alle in den vorherigen Schritten generierten Stammbäume unter Berücksichtigung eines spezifischen Erbmechanismus generiert sind, lässt sich eine Vielzahl von ihnen ohne weiterführende Informationen keinem Vererbungsmechanismus eindeutig zuordnen, weil der Stammbaum keine Schlüsselstellen oder sonstige Charakteristika aufweist, die eine eindeutige Identifikation des Vererbungsmechanismus zulassen. Es erscheinen dann

mehrere Mechanismen mehr oder weniger gleichermaßen wahrscheinlich. Um diesem Problem zu begegnen haben wir ein Analysemodul entwickelt, welches in einem dritten Schritt überprüft, ob sich ein generierter Stammbaum hinreichend eindeutig einem Vererbungsmechanismus zuordnen lässt. Dieses Analysemodul nutzt ausschließlich die Informationen, die auch den Studierenden zur Verfügung stehen. Dieser Prüfungsschritt wäre vermutlich prinzipiell vermeidbar, wenn bereits während der Generierung des Stammbaums eindeutige Identifikationsmerkmale innerhalb des Stammbaums erzwungen würden. Dieser Ansatz wurde allerdings zum einen aus Komplexitätsgründen und zum anderen, um die realistische Verteilung der Allele über die Generationen hinweg so wenig wie möglich durch zusätzliche Konstruktionsregeln zu manipulieren, und so potentiell die Struktur der erzeugten Stammbäume systematisch zu beeinflussen, verworfen.

Das Analysemodul erkennt einen Stammbaum dann als eindeutig bestimmbar, wenn sich alle alternativen Vererbungsmechanismen mit hinreichender Sicherheit ausschließen lassen. Beispielhaft lässt sich dies anhand des in Abbildung 2 dargestellten Stammbaums erläutern. Die Person mit der Nummer 120 besitzt ein Merkmal, welches seine Eltern nicht besitzen. Dementsprechend wird dieses Merkmal nicht dominant, sondern rezessiv vererbt. Da mit Person 106 auch eine Frau Merkmalsträgerin ist, kann das merkmalscodierende Gen außerdem nicht auf dem Y-Chromosom liegen (da Frauen kein Y-Chromosom besitzen). Auch eine Lokalisierung des rezessiven Gens auf dem X-Chromosom lässt sich ausschließen, da Mann 112 andernfalls Merkmalsträger sein müsste. Auf Grundlage dieser Überlegungen lässt sich der Stammbaum eindeutig bestimmen; dargestellt ist ein autosomal-rezessiv vererbtes Merkmal.

Da der Anteil eindeutig bestimmbarer Stammbäume je nach Erbgang sehr gering ist, haben wir vor Beginn des Einsatzes in der Lehrveranstaltung einen Aufgabenpool bestehend aus 150 Aufgaben (jeweils zehn für einen von fünf Vererbungsmechanismen und eine von drei Größen) generiert. Hierzu wurden mithilfe der zuvor vorgestellten Schritte 6771 Stammbäume generiert und analysiert, wobei für jeden eindeutig bestimmbaren Stammbaum lediglich ein erzeugungsrelevanter *seed key*, der Vererbungsmechanismus und die gewünschte Generationenzahl in einem DataFrame gespeichert werden müssen.

Die tatsächliche Generierung der Aufgabe findet dann erst statt, wenn ein Studierender eine Stammbaumaufgabe anfordert. Aus dem zuvor genannten DataFrame wird zufallsbasiert eine Zeile mit Konstruktionsvariablen ausgewählt und auf dieser Basis die entsprechende Aufgabe generiert, wobei erst jetzt die entsprechende Abbildung erzeugt wird. Ergänzend wird auch das Analyseskript erneut aufgerufen, da es nicht nur den vorliegenden Erbgang bestimmt, sondern darüber hinaus individuelles Feedback generiert, indem es Personen innerhalb des Stammbaums identifiziert, anhand derer sich die spezifischen, nicht zutreffenden Vererbungsmechanismen ausschließen lassen. Dieses Feedback wird den Studierenden während der Bearbeitung immer dann angezeigt, wenn sie sich für einen falschen Vererbungsmechanismus entscheiden. Sie bekommen dann die

Gelegenheit, sich für einen alternativen Vererbungsmechanismus zu entscheiden, können also mehrfach ihre Antwort überprüfen.

Bei der Auswahl der Aufgaben wurden, wie angestrebt, alle fünf relevanten Vererbungsmechanismen ($N_{min}$ = 246; $N_{max}$ = 297) und die drei möglichen Stammbaumgrößen ($N_{min}$ = 407; $N_{max}$ = 463) in vergleichbarer Häufigkeit generiert. Insgesamt wurden 1304 Aufgabenbearbeitungen von 134 Studierenden vollzogen, was im Durchschnitt 9,7 Bearbeitungen pro Studierendem entspricht, wobei die individuelle Nutzungshäufigkeit stark variiert (61 Personen mit bis zu 3 Bearbeitungen, maximale Nutzungshäufigkeit einer Person = 119). Die Stammbaumaufgaben besitzen eine angemessene Schwierigkeit für die Studierenden, in 65,9 % der Fälle wurde im ersten Versuch die richtige Lösung ausgewählt. Die Wahrscheinlichkeit innerhalb von zwei oder drei Versuchen die richtige Lösung auszuwählen belief sich kumuliert auf 84,7 % beziehungsweise 92,2 %.

Auf Grundlage der bisherigen Ergebnisse bewerten wir die Aufgabengenerierung und -selektion als zufriedenstellend, da Stammbäume in allen Größen und zu allen Erbgängen in ausreichender Zahl generiert und bearbeitet wurden. Die durchschnittliche Nutzungshäufigkeit der Studierenden legt nahe, dass die Variabilität der generierten Aufgaben so groß ist, dass das Analysieren von Stammbäumen mithilfe dieser Aufgabe mehrfach geübt werden kann. Darüber hinaus wird die Qualität der Aufgaben zukünftig anhand von Begründungen evaluiert, die die Studierenden bei der Bearbeitung zusätzlich angeben müssen. Außerdem ist eine weiterführende Wirksamkeitsanalyse auf Grundlage von Klausurergebnissen geplant, wobei das Studiendesign und die hohe Variabilität der Aufgaben besondere Herausforderungen darstellen.

## 6 Analyseergebnisse

Als positive Erkenntnis der beiden unabhängigen Fallbeispiele kann festgehalten werden, dass es jeweils möglich war, die fachliche Komplexität der Aufgabenstellung soweit in algorithmischer Form beziehungsweise in Form von Aufgabenkonfigurationen festzuhalten, dass die Erstellung fachlich sinnvoller Aufgabeninhalte vollständig automatisiert erfolgen konnte. Ferner wurde festgestellt, dass die so erzeugten Aufgaben auf Basis ihrer Lösungswahrscheinlichkeit als ausreichend schwierig zu bewerten sind, so dass die Automatisierung nicht mit einer verminderten Aufgabenqualität einhergeht.

Gleichzeitig konnte jedoch auch festgestellt werden, dass die automatische Erzeugung einer Nachsteuerung bedarf, die bei den botanischen Aufgaben aufgrund der großen Menge an erzeugten Distraktoren größer ist als bei den genetischen Stammbäumen.

## 7 Fazit und kommende Arbeiten

Der vorliegende Beitrag zeigt, dass fachlich hochspezifische Aufgabentypen der Biologie erfolgreich automatisch generiert werden können. Das dazu nötige Fachwissen kann im Wesentlichen im Rahmen des Konstruktionsprozesses codiert werden, um eine hohe Item-Qualität zu erzielen, während eine Nachsteuerung auf Basis statistischer Merkmale mit geringerem Fachwissen möglich erscheint.

Neben der iterativen Verbesserung der vorhandenen Aufgaben und der Erweiterung der Aufgabenpools innerhalb der jeweiligen Disziplin erscheint es möglich, das Aufgabenformat aus dem Bereich der Botanik auf weitere Kontexte der Biologie zu übertragen, welche ebenfalls eine formelbasierte Darstellung umfassen. Hier bietet sich zum Beispiel das Erlernen von Zahnformeln im Bereich der Zoologie an. Ferner könnte ein abgeleitetes Format beim Erlernen von anatomischen Begriffen genutzt werden, die sich häufig aus verschiedenen Begriffen zusammensetzen (z.B. *Musculus adductor magnus* und *Musculus adductor longus*).

Die Aufgabe zu Familienstammbäumen und das zugrundeliegende Skript lässt sich zwar nicht ohne weiteres zur Entwicklung anderer biologiespezifischer Aufgaben nutzen, verdeutlicht aber, dass es möglich ist, auf Grundlage fachspezifischer Diagramme automatisiert Problemlöseaufgaben zu generieren, was beispielsweise im Fachbereich der Evolution mit Dendrogrammen (im Sinne phylogenetischer Stammbäume) denkbar wäre.

**Danksagung:** Wir danken den Kolleginnen und Kollegen aus der Botanik und Genetik für die Unterstützung bei der Aufgabenentwicklung und die Möglichkeit, diese in den Lehrveranstaltungen einzusetzen. Die Forschungsarbeit in diesem Paper wurden vom BMBF unter Förderkennzeichen 01PL16075 gefördert.

Literaturverzeichnis

[Al12] Alsubait, T. et al.: Next generation of e-assessment: automatic generation of questions. International Journal of Technology Enhanced Learning, 4(3/4), S. 156-171, 2012.

[Al15] Alsubait, T.: Ontology-based multiple-choice question generation. PhD thesis, School of Computer Science, The University of Manchester, 2015.

[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.

[Be08] Bennett, R. L. et al.: Standardized human pedigree nomenclature. Update and assessment of the recommendations of the National Society of Genetic Counselors. Journal of genetic counseling 17(5), S. 424–433, 2008.

[Co17] The Concord Consortium: Teaching Genetics with Dragons - The Concord Consortium's Genetics Education Software. https://concord.org/teaching-genetics/dragons, 2017.

[Da13] Daroczi, G.. The sandboxR Package: Filtering “Malicious” Calls in R. https://github.com/Rapporter/sandboxR, 2013.

[GLZ18] Gierl, M., Lai, H., Zhang, X.: Automatic Item Generation. In M. Khosrow-Pour, D.B.A. (Hrsg.): Encyclopedia of Information Science and Technology, Fourth Edition. Hershey, PA: IGI Global, S. 2369-2379, 2018.

[Gr15] Graw, J.: Genetik. 6. Aufl. Berlin, Heidelberg: Springer, 2015.

[Ha94] Hackling, M. W.: Application of genetics knowledge to the solution of pedigree problems. Research in Science Education 24(1), S. 147–155. 1994.

[KMK04] Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (Hrsg.): Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie. Beschluss der Kultusministerkonferenz vom 01.12.1989 i.d.F. vom 05.02.2004, 2004.

[KMK05] Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (Hrsg.): Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004. München: Luchterhand, 2005.

[Ku15] Kurt-Karaoglu, F.: A Framework for Generic Exercises with Mathematical Content. In: 2015 International Conference on Learning and Teaching in Computing and Engineering, IEEE, S. 70-75, 2015.

[La16] Lai, H. et al.: Using Automatic Item Generation to Improve the Quality of MCQ Distractors. Teaching and Learning in Medicine, 28, Routledge, S. 166-173, 2016.

[Ma00] Mayr, E.: Das ist Biologie, Die Wissenschaft des Lebens, Spektrum, Heidelberg, 2000.

[Me13] Means, B. et al.: The Effectiveness of Online and Blended Learning: A Meta-Analysis of the Empirical Literature. Teachers College Record 115(3), S. 1-47, 2013.

[Ot17] Otto, B. et al.: Evaluation einer Statistiklehrveranstaltung mit dem JACK R-Modul. In: DeLFI 2017 – Die 15. e-Learning Fachtagung Informatik, S. 75-86, 2017.

[Sa13] Sangwin, C.: Computer aided assessment of mathematics. OUP Oxford, 2013.

[SB14] Škopljanac-Mačina, F.; Blašković, B.: Formal concept analysis–overview and applications. Procedia Engineering, 69, S. 1258-1267, 2014.

[SNT17] Schmiemann, P.; Nehm, R. H.; Tornabene, R. E.: Assessment of Genetics Understanding. Science & Education, 26(10), S. 1161–1191, 2017.

[STS14] Sinnwell, J. P.; Therneau, T. M.; Schaid, D. J.: The kinship2 R package for pedigree data. Human heredity 78(2), S. 91–93, 2014.

[TG11] Teufel, A.; Galante, T.: Humangenetik (Basics, 1. Aufl.). Elsevier, 2011.

[TT10] Tsui, C. Y., & Treagust, D. Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. International Journal of Science Education, 32(8), 1073-1098. 2010

[Ur03] Urbanek, S.: Rserve - a fast way to provide R functionality to applications. In: Proc. of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 2003.

# Ein Format für Bewertungsvorschriften in automatisiert bewertbaren Programmieraufgaben

Robert Garmann[1]

**Abstract:** Automatisiert bewertbare Programmieraufgaben definieren Tests, die auf Einreichungen angewendet werden. Da Testergebnisse nicht mit Bewertungsergebnissen gleichzusetzen sind, schlagen wir ein Beschreibungsformat vor, das Testergebnisse auf Bewertungsergebnisse abbildet. Lehrkräfte können die Abbildungsvorschrift an ihren Lehrkontext anpassen. Der Vorschlag ist unabhängig von den beteiligten Autobewertern, von den eingesetzten Benutzungsschnittstellen und von der zu lernenden Programmiersprache einsetzbar. Das Format basiert auf verschachtelten Bewertungskategorien, welche um ein Nullifikationen-Konzept ergänzt werden.

**Keywords:** Automatisierte Bewertung, Programmieraufgaben, Grader, Autobewerter, E-Assessment, Bewertungsschema, ProFormA, Softwaretest

## 1 Einleitung

Automatisiert bewertbare Programmieraufgaben werden häufig im formativen E-Assessment von Informatik- und verwandten Lehrveranstaltungen mit großen Teilnehmerzahlen eingesetzt, um das regelmäßige Programmieren zu fördern. Abb. 1 gibt ein Beispiel.

*Aufgabe: Schreiben Sie eine Java-Klasse `Circle` mit zwei statischen Methoden `calcArea` und `calcRadius`. Die erste Methode soll für einen gegebenen Radius die Kreisfläche berechnen, die zweite Methode soll die Umkehrfunktion berechnen.*

| | | | | |
|---|---|---|---|---|
| Syntax (10%) | | | ✓ | 10% |
| Funktion (60%) | | | | 18% |
| | calcArea (×30%) | | ✓ | |
| | calcRadius (×70%) | | ✗ | |
| Wartbarkeit (30%) | | | | 0% |
| | Stil | Anweisungen/Zeile (×30%) | (✓) | |
| | | Bezeichner | ✗ | |
| | Kommentare (×70%) | | ✗ | |
| Summe | | | | 28% |

Abb. 1: Beispielaufgabe (links) und Bewertungsergebnis (rechts)

Um Anforderungen an ein Format für Bewertungsvorschriften zu motivieren, stellen wir in Tab. 1 ein mögliches Bewertungsschema vor, das von menschlichen Bewertern eingesetzt werden könnte. Die in Abb. 1 dargestellte, für eine konkrete Einreichung vorgenommene Bewertung könnte in ähnlicher Weise automatisiert durch einen sog. Autobewerter (Grader) erfolgen. Solche Bewertungsschemata haben, wenn sie vorab veröffentlicht werden, im summativen Assessment das Potenzial, möglichst transparent und möglichst exakt diejenigen Lernanreize zu setzen, die von der Lehrkraft intendiert werden [Bi99]. Zudem gibt es Hinweise darauf, dass transparente Bewertungskriterien auch im formativen Assessment die Lernleistung von Studierenden verbessern können [PJ13].

[1] Hochschule Hannover, Fakultät IV – Wirtschaft und Informatik, Ricklinger Stadtweg 120, 30459 Hannover, robert.garmann@hs-hannover.de

doi:10.18420/delfi2019_73

Aufwändig entwickelte Aufgaben werden möglichst häufig von verschiedenen Lehrkräften wiederverwendet. Das ProFormA-Aufgabenformat [St15] vereinfacht den Austausch zwischen Lehrenden und die Wiederverwendung von Aufgaben über System- und Institutsgrenzen hinweg. Eine automatisiert bewertbare Aufgabe besteht u. a. aus dem Aufgabentext, den sog. *Tests* zur automatischen Prüfung einer Einreichung sowie der *Bewertungsvorschrift* (*grading-hints*), die aus Testergebnissen ein Bewertungsergebnis berechnet. Während Tests in der Regel sehr detaillierte Angaben enthalten, sind Bewertungsvorschriften häufig (und so auch in Version 1.1 des ProFormA-Formats) knappgehalten, wenn sie nicht gar vollständig fehlen. Bei diesem Mangel können Aufgabenautoren und Lehrkräfte keine Angaben in der Aufgabe hinterlegen, wie die einzelnen zu bewertenden Aspekte zu gewichten sind, so dass ein Autobewerter die Testergebnisse eins-zu-eins als Bewertungsergebnisse ausgibt. Dieser Beitrag befasst sich mit der Frage, wie diese Lücke zu schließen ist. Da es viele verschiedene Auffassungen gibt, welcher Aspekt wie zu gewichten ist, soll diesbezüglich eine gewisse Offenheit des Formats bestehen.

| **Dim.** | **1,0** | **0,7** | **0,3** | **0,0** |
|---|---|---|---|---|
| **Syntax, 10%** | Fehlerfrei | | | Fehlerhaft. Oder: Trivialabgabe[a]. |
| **Funktion, 60%** | Zwei Methoden funktionieren | Nur calcRadius funktioniert. | Nur calcArea funktioniert. | Nichts funktioniert |
| **Wartbarkeit, 30%** | Kommentare *vorhanden* und Code ist stil*konform*[b] | Kommentare *vorhanden,* aber Code *verletzt* Stilregeln[b] | Kommentare *fehlen*, aber Code ist stil*konform*[b] | Kommentare *fehlen* und Code *verletzt* Stilregeln[b]. Oder: Trivialabgabe[a]. |

[a] eine Trivialabgabe (bspw. Klasse ohne Methoden) gilt als Versagen in Syntax und Wartbarkeit (und in Funktion, aber das muss ja nicht extra erwähnt werden).
[b] Stilkonformität: max. eine Anweisung pro Zeile *und* konventionsgemäße Bezeichner

Tab. 1: Bewertungsschema für die Beispielaufgabe

## 2 Methode

Unser Forschungsansatz ist dem Design-Based Research zuzuordnen. Die auf Vorarbeiten in [Ga15, Ga16] basierenden, zunächst im kleinen Rahmen für ausgesuchte Lehrveranstaltungen entwickelten Werkzeuge und Methoden wurden von mir und weiteren Lehrenden selbst im Lehreinsatz erprobt und anschließend im Rahmen regelmäßiger Interventionen evaluiert und überarbeitet. Neben dem Ziel, praxistaugliche Systeme und Methoden für konkrete Lehre vorgegebener Programmiersprachen zu erreichen, werden Erkenntnisse von größerer Allgemeingültigkeit angestrebt, die sich auf weitere zur Autobewertung eingesetzte Systemlandschaften und Programmiersprachen übertragen lassen. Forschungsziel ist, die in Abschnitt 1 erwähnte Lücke für automatisch bewertete Programmieraufgaben zu schließen. Manuelle Bewertungen oder Bewertungen anderer Aufgabentypen stehen nicht im Fokus.

Um sich dem Ziel der Verallgemeinerbarkeit zu nähern, wurde zum einen die vergleichsweise etablierte Spezifikation IMS QTI zum Austausch von Tests in der aktuellen Version 2.2 konsultiert[2], ob die darin spezifizierten Datenmodelle verwendet werden können. Des Weiteren wurden die Bewertungsmechanismen bestehender Grader und LMS untersucht. Die Auswahl umfasste diejenigen Systeme, die von Lehrenden und Forschenden der ProFormA-Arbeitsgruppe des eCult+-Projekts[3] eingesetzt werden, deren Angehörige allesamt intensiv an der Entwicklung und Nutzung von Systemen und Methoden für die automatisierte Programmbewertung arbeiten. Untersucht wurden die LMS Moodle und LON-CAPA sowie die Grader Praktomat, JACK, Graja, aSQLg, GATE, ASB, PABS und VEA, die allesamt bspw. in [Bo17] beschrieben sind. Die aus den Interventionen entstandenen Weiterentwicklungen wurden kontinuierlich in der Arbeitsgruppe vorgestellt und diskutiert. In den Diskussionen entstanden Anforderungen an und Lösungsvorschläge für ein Format für Bewertungsvorschriften, welches Anspruch auf größere Allgemeingültigkeit erheben kann und Teil des ProFormA-Formats werden kann.

Den vorliegenden Beitrag beginnen wir mit einem eingehenderen Blick auf die Eigenarten von Tests und einer Einordnung des Qualitätsbegriffs von Bewertungsschemata. Abschnitt 4 beschreibt die im o. g. Entwicklungsprozess entstandenen Anforderungen, vor deren Hintergrund wir ab Abschnitt 5 die entstandenen Lösungsvorschläge herleiten und argumentativ bewerten. Das Beispiel aus Abb. 1 und weitere, aus der konkreten Lehrpraxis erwachsende Normalfälle und Sonderfälle dienen dabei zur Illustration und Evaluation. Quantitative Analysen z. B. durch Umfragen unter Studierenden und Lehrenden wurden nicht durchgeführt. Schließlich beschreibt Abschnitt 7 eine weitergehende Evaluation durch eine konkrete Implementierung. Weitere Beispiele und Untersuchungen sind in einem Forschungsbericht dokumentiert [Ga19].

## 3 Tests und Bewertungen

Typische Tests zur automatischen Prüfung einer Einreichung reichen von der einfachen Compilierung über dynamische, in einem Unittest-Framework verfasste Tests bis hin zu statischen Codeanalysen und Effizienzprüfungen. Weitere Tests sind etwa Plagiat-Checks, aber auch einfache Prüfungen auf Überschreitungen von Abgabeversuchen und -terminen. Es hat sich für die Ableitung von Bewertungsvorschriften als nützlich herausgestellt, zwei Klassen von Tests zu bilden (s. Tab. 2). Unter robuste Tests (R-Tests) fällt die dynamische Funktionsprüfung, die sich bspw. als Unittest oder als Prüfung der Ergebnisse von SQL-Abfragen manifestiert, aber auch die statische Analyse des Quelltextes auf Anwesenheit erwünschter Konstrukte. R-Tests prüfen Eigenschaften, welche nur durch konstruktive Maßnahmen eines Studierenden erreichbar sind, der die zu prüfenden Kompetenzen besitzt[4]. Dagegen können fragile Tests (F-Tests) für sich genommen z. B. durch funktionslose Trivialabgaben bestanden werden, ohne dass der Studie-

[2] http://www.imsglobal.org/question
[3] http://ecult.me
[4] Ob dies der einreichende Studierende ist oder eine Plagiatquelle, ist dabei unerheblich.

rende alle zu prüfenden Kompetenzen besitzt. Zu den F-Tests zählen wir Compilierung, Regelverletzungen prüfende statische Codeanalysen, Plagiatstests, Effizienzprüfer, etc. Alle Tests bewerten eine Einreichung mit einem Ergebnis in [0,1], wobei 1 einen vollständigen Erfolg benennt und 0 einen vollständigen Misserfolg.

| | |
|---|---|
| **(R)obust** | Prüft Merkmale, die nur durch Kompetenz erreichbar sind. |
| **(F)ragile** | Prüft Merkmale, die isoliert gesehen mit Trivialabgaben erreichbar sind. |

Tab. 2: Testtypen

Offenbar ist es sinnvoll, Einreichungen immer mit beiden Testarten zu testen. Ein Autobewerter, der nur auf F-Tests setzt, ist leicht durch eine Trivialabgabe wie einen funktionslosen Programmrahmen auszuhebeln. Häufig reicht es auch nicht, ausschließlich R-Tests einzusetzen, wenn von Studierenden die Einhaltung von Randbedingungen gefordert wird (Effizienz, Wartbarkeit, akademischer Ehrenkodex, etc.), deren Prüfung mit R-Tests schwierig ist. Automatisiert bewertete Aufgaben enthalten daher üblicherweise mehrere Tests, deren Ergebnisse zu Bewertungsergebnissen aggregiert werden.

Häufig nutzen Menschen sog. „rubrics“ (Bewertungsschemata), um den Bewertungsprozess zu strukturieren (vgl. Tab. 1). Diese können – wie im Beispiel dargestellt – aufgabenübergreifende und aufgabenspezifische Kriterien enthalten. Aus den je Zeile ermittelten Einzelbewertungen wird dann mit Hilfe einer sog. *Aggregierungsstrategie* eine Gesamtbewertung errechnet. Häufig lassen sich die Bewertungsdimensionen nicht vollkommen unabhängig voneinander gestalten [Sa89]. Wenn sich Dimensionen überlappen oder Teile von gröberen Dimensionen sind, können Leistungen in der einen Dimension ggf. nur bei gleichzeitigen Leistungen in einer anderen Dimension erbracht werden. Eine Gesamtbewertung einfach als Summe von Teilergebnissen zu berechnen wird der Abhängigkeit der Dimensionen ggf. nicht gerecht.

Die Auswahl der Bewertungsdimensionen, die Gestaltung der Bewertungsstufen sowie die Konstruktion der Gesamtformel haben Auswirkungen auf die Qualität eines Bewertungsschemas, d. h. darauf, dass alle relevanten und keine irrelevanten Bewertungsaspekte Beachtung finden und dass es keine unerwünschten Variationen zwischen Bewertungen eines Bewerters (intra-rater) und zwischen Bewertungen verschiedener Bewerter (inter-rater) gibt [JS07]. Im automatisierten E-Assessment sind intra-rater-Variationen nicht zu erwarten, wenn wir davon ausgehen, dass der Autobewerter ausschließlich deterministische Testmethoden oder erwartungskonforme probabilistische Testmethoden nutzt. Unter inter-rater-Variationen verstehen wir im E-Assessment hauptsächlich die Diskrepanz zwischen dem Ergebnis des Autobewerters und dem gleichzeitig für dieselbe Einreichung erstellten Ergebnis eines menschlichen Bewerters.

In diesem Beitrag schlagen wir weder ein konkretes Bewertungsschema vor noch bewerten wir dessen inter-rater-Qualität. Ziel dieses Beitrags ist auf einer Meta-Ebene ein Format zur Beschreibung von Bewertungsschemata, das durch seine Betonung der Nachvollziehbarkeit gute inter-rater-Qualitäten erwarten lassen kann.

# 4 Anforderungen

Tab. 3 listet die Anforderungen an ein Format für Bewertungsvorschriften auf, die in dem in Abschnitt 2 beschriebenen Prozess enstanden sind. R1 nennt den Status quo, dass viele LMS zu einer Aufgabe ein metrisch oder ordinal skaliertes Zahlergebnis speichern. Die dazu notwendige Aggregierung mehrerer, dem Intervall [0,1] entstammender Teilergebnisse setzt aus testtheoretischer Sicht viel Wissen über die Natur der Teilergebnisse voraus. Es besteht jedoch explizit keine Anforderung an die Berücksichtigung der Testtheorie. Wenn ein Aufgabenautor eine gewichtete Summe für nicht intervallskalierte Teilergebnisse bilden will, dann soll er das dürfen.

| | |
|---|---|
| R1 | Die Bewertungsvorschrift soll die Errechnung eines skalaren Gesamtergebnisses für eine Aufgabe ermöglichen. |
| R2 | Besonderheiten einzelner Systeme und Programmiersprachen sollen nicht im Format implementiert sein. |
| R3 | Die Bewertungsvorschriften sind für eine Lehrkraft leicht verständlich und an den Lehrkontext anpassbar. Eine Skalierung mit einem Faktor ist möglich. |
| R4 | Die Bewertungsvorschrift muss aus Sicht der Lehrkraft ermöglichen, voneinander unabhängige Bewertungsaspekte getrennt „zu würdigen“. |
| R5 | Häufig wiederkehrende Teilaspekte der Bewertung und deren Verknüpfung sollen durch die Lehrkraft und durch etwaige maschinelle Unterstützung des LMS einfach auf mehrere Aufgaben einer Lehrveranstaltung „ausgerollt“ werden können. |
| R6 | Das maschinell erstellte Bewertungsergebnis soll für Studierende verständlich und nachvollziehbar sein. |
| R7 | Die Bewertungsvorschrift soll Eigenheiten der von Testwerkzeugen generierten Testergebnisse berücksichtigen. |

Tab. 3: Anforderungen

Anforderungen R2 und R3 sollen die Wiederverwendung von Aufgaben fördern. Der in der o. g. Arbeitsgruppe häufig von Lehrkräften geäußerte Wunsch, die maximal erreichbare Punktzahl je Aufgabe anpassen zu können, fordert eine einfache Skalierung.

Im formativen E-Assessment von Programmieraufgaben werden häufig beliebig viele Abgabeversuche bis zum Abgabetermin gewährt, die unmittelbar mit automatisiertem Feedback beantwortet werden. Manche Lehrende beobachten das unerwünschte Studierendenverhalten, Abgaben in hoher Frequenz einzureichen und dabei immer nur leicht zu verändern, bis (zufällig?) eine Punktzahl erreicht ist, die einigermaßen zufrieden stellt. Dem entgegenwirkend lassen transparente, jeden einzelnen Aspekt würdigenden Bewertungskriterien (Anforderung R4) erwarten, dass Studierende vor einer Einreichung anhand des Bewertungsschemas prüfen, ob sie alle Bewertungsaspekte berücksichtigt haben [PJ13]. Um den Aufwand bei vielen Aufgaben beherrschbar für die Lehrkraft zu gestalten, wurde R5 aufgenommen. Man soll etwa zentral für alle Aufgaben eines Semesters festlegen können, mit welchen Prozentsätzen Syntax, Funktion und Wartbarkeit in das Gesamtergebnis eingehen. Werden allerdings neben klassischen Programmieraufga-

ben, bei denen Studierende selbst programmieren müssen, auch sog. code execution exercises gestellt, bei denen Studierende die Ausgabe eines gegebenen Programms nennen müssen, kann R5 nur bedingt erfüllt werden, da der letztgenannte Aufgabentyp keine Bewertungsaspekte für Syntax und Wartbarkeit kennt. R5 fordert letztendlich, dass aufgabenübergreifende Bewertungsaspekte für LMS identifizierbar vorliegen müssen.

Im E-Assessment überbringt eine Maschine Feedback. Dabei gewinnt die in [PJ13] genannte, durch nachvollziehbare Bewertungskriterien geförderte Selbstregulation des Lernens durch Angstreduktion an Bedeutung. Anforderung R6 will bei Studierenden kein lernbehinderndes Gefühl aufkommen lassen, Opfer von willkürlicher Bewertung zu sein.

Testwerkzeuge wurden für die Softwarenentwicklung und nicht für die Benotung von Lernleistungen entwickelt (vgl. Abschnitt 3). Anforderung R7 wurzelt in der Erkenntnis, dass Testergebnisse nicht mit Bewertungsergebnissen verwechselt werden dürfen.

## 5 Aggregierungsstrategien

Es lassen sich bei den in Abschnitt 2 genannten Standards, Gradern und LMS verschiedene Aggregierungsstrategien beobachten (s. Tab. 4). In diesem Abschnitt wollen wir hieraus die den Anforderungen genügenden Strategien auswählen.

Verbreitet werden gewichtete Summen eingesetzt, bspw. bei voneinander weitgehend unabhängigen R-Tests, die je einen Funktionsaspekt prüfen. Die Min.-Strategie kommt z. B. bei zwei Stilregeln (F-Tests) zum Einsatz, wobei je nach dem vom Studenten gewählten Lösungsansatz mal die eine und mal die andere Stilregel verletzt sein kann, aber nicht beide. Eine Aufgabe mit einer Max.-Strategie fordert bspw. entweder den Einsatz der for-Schleife oder der while-Schleife und verknüpft dabei zwei zugehörige R-Tests. Vor dem Hintergrund von R1, R3, R6 und R7 sind alle drei Strategien (Summe, Min., Max.) sinnvoll und notwendig. Zu R3 sei bemerkt, dass Studierende häufig fragen, welchen Einfluss ein bestimmter Bewertungsaspekt (bspw. der Aspekt *calcRadius* in Abb. 1) auf die Gesamtpunktzahl einer Aufgabe hat. Mit den Strategien Summe, Min. und Max. lässt sich die Antwort (hier *Funktion*□*calcRadius*=60%□70%=42%, vgl. Abb. 1) leicht auf eine von der Lehrkraft vorgegebene neue Gesamtpunktzahl skalieren.

Für die Conditional-Aggregation beschreiben wir ein Beispiel, an dem das Ergebnis F eines robusten Funktionstests und die Ergebnisse $P_1$ und $P_2$ zweier fragiler Plagiatstests beteiligt sind. Die Plagiatstests liefern unabhängig voneinander je einen Erfolgswert zwischen 0 (sicheres Plagiat) und 1 (sicher kein Plagiat). Wenn mindestens eines der beiden Plagiatstestergebnisse 1 lautet, soll im Zweifel für den Studenten angenommen werden, dass kein Plagiat vorliegt. Das kombinierte Ergebnis ist dann gleich F. Ansonsten soll vollständiger Misserfolg wegen Vorliegen eines Plagiats angenommen werden:

$$\text{Gesamterfolg} = \begin{cases} \text{F, falls } P_1 = 1 \text{ oder } P_2 = 1 \\ 0, \text{ sonst} \end{cases}$$

Unter der Voraussetzung, dass $P_1$ und $P_2$ Ergebnisse aus {0,1} liefern, ließe sich der Gesamterfolg auch als eine Kombination von Min. und Max. darstellen [Ga19], vermutlich jedoch zu Lasten der intuitiven Verständlichkeit (R6).

| | |
|---|---|
| **Summe** | Je mehr Einzeltests erfolgreich verlaufen, desto höher wird der Gesamterfolg gewertet (Summe der Einzelerfolge) |
| **Min. (worst)** | Der Gesamterfolg ist der geringste Erfolg aller Einzeltests |
| **Max. (best)** | Der Gesamterfolg ist der größte Erfolg aller Einzeltests |
| **Conditional** | Bedingungsgesteuerte Verknüpfung von Einzelerfolgen zu einem Gesamterfolg |
| **Assoziativ-speicher** | Mit menschlichen Bewertungen trainierter Assoziativspeicher, dessen aus den Einzelerfolgen berechneter Gesamterfolg möglichst gut mit Ergebnissen menschlicher Bewerter übereinstimmt (inter-rater-Variation). |
| **Spreadsheet** | Flexible Kombination von Einzelerfolgen durch eine Tabelle, deren Zellen Funktionen einer Tabellenkalkulation enthalten. |
| **Median** | Der Gesamterfolg ist der mittlere Erfolg aller Einzeltests |
| **Vektor** | Der Gesamterfolg ist ein Tupel der Erfolge der Einzeltests |

Tab. 4: Aggregation von Testergebnissen

Zur Eignung von Assoziativspeichern sei als Gegenbeispiel das Anwendungsgebiet des *Automatic essay scoring (AES)* genannt, das dem automatisierten Bewerten von Programmen, die als frei innerhalb syntaktischer Regeln zu formulierende Aufsätze verstanden werden können, ähnlich ist. AES hat eine lange Tradition in der Definition hunderter Bewertungskriterien, die teilweise in kommerziellen Werkzeugen als Geschäftsgeheimnis implementiert sind. Aus diesen Kriterien wird mit Hilfe von Trainingsdaten und mit statistischen Methoden bzw. mit Methoden der Künstlichen Intelligenz eine Aggregierungsstrategie berechnet. Die daraus errechneten Bewertungsergebnisse entbehren häufig jeder Nachvollziehbarkeit (R6). Dies wird bspw. in [AB06] zum Anlass genommen, die Bewertung im AES-System e-rater auf ein vollständig transparentes Bewertungsschema mit wenigen, nachvollziehbaren Bewertungskriterien umzustellen.

Ein Spreadsheet bietet zwar volle Transparenz über den Bewertungsmechanismus. Allerdings kann ein Spreadsheet sehr komplexe Berechnungen vornehmen. Bspw. könnte ein Testteilergebnis komplexen mathematischen Funktionen unterworfen werden, bevor es mit anderen Testteilergebnissen akkumuliert wird. Es ist dann zumindest mühsam, zu erkennen, welchen Einfluss ein einzelnes Testteilergebnis auf das Gesamtergebnis hat. Die Anforderungen R3 und R6 sind nur bedingt erfüllt.

Der Median ist zwar das testtheoretisch geeignete Mittel, um oridinalskalierte Teilergebnisse zu aggregieren. Vor dem Hintergrund von R6 entscheiden wir uns gegen die Einbindung des Medians in der Annahme, dass dessen Kenntnis unter Studierenden nur gering verbreitet ist.

Schließlich bemerken wir, dass ein Vektor als Gesamtergebnis R1 verletzt. Vorteil einer Lösung, in der der Grader alle Einzelergebnisse an das LMS in Form eines Vektors lie-

fert, wäre, dass das LMS die Teilergebnisse auf sehr flexible und aufgabenübergreifende Weise zu einer Kursnote verknüpfen könnte. Nachteil einer solchen Lösung wäre, dass bei weitem nicht jedes LMS solch weitreichende Bewertungsfunktionen unterstützt (R2).

Zusammenfassend sollen gewichtete Varianten von Summe, Min. und Max. sowie die Conditional-Strategie in den Formatvorschlag aufgenommen werden.

IMS QTI kann prinzipiell durch individuelle, automatisiert bewertbare Programmieraufgaben mittels *custom response processing* erweitert werden. Für solch individuelle Aufgaben sieht QTI jedoch keine standardisierten Bewertungsvorschriften vor. QTI definiert auf der nächsthöheren Ebene in sog. *Assessment*s mit der *ExpressionGroup* ein großes Arsenal von Verrechnungsregeln zur Verknüpfung mehrerer Aufgaben. Hierunter sind auch die im vorstehenden Absatz selektierten Aggregierungsstrategien. Hauptvorteil bei der Verwendung eines Standards oder eines etablierten Datenformats ist, dass es bereits LMS gibt, die das Format unterstützen[5]. Der naheliegende Ansatz, eine Teilmenge des QTI-Datenmodells zu nutzen, um Bewertungsvorschriften innerhalb einer ProFormA-Aufgabe zu definieren, wurde verworfen, weil bestehende LMS das Datenformat eben nicht für eine einzelne Aufgabe unterstützen, sondern für mehrere Aufgaben, die als Teile eines größeren Quizzes oder Assessments verknüpft werden.

## 6 Formatspezifikation und Illustration am Beispiel

Um Ergebnisse fragiler und robuster Tests separat aggregieren zu können (R7), reicht eine einstufige Hierarchie nicht aus. Mehrstufige Hierarchien ineinander verschachtelter Bewertungskategorien werden in verschiedenen LMS und Gradern bereits genutzt, so dass wir diese Struktur in das Format übernehmen wollen (vgl. Abb. 2 links). Die Hierarchie*node*s erhalten *id*s und Aggregierungs*function*s[6]. Anforderung R5 wird gut erfüllt, wenn Hierarchieknoten auf oberen Stufen *id*s besitzen, die sich alle Aufgaben einer Lehrveranstaltung teilen. Dann kann das LMS die gemeinsamen Teile der Bewertungsvorschriften aller Aufgaben einer Lehrveranstaltung automatisiert harmonisieren.

Eine Baumstruktur hat den Vorteil einer vorgegebenen Lesestruktur, die durch in den Baumknoten enthaltene *title*s inhaltlich strukturiert wird. Studierende können einen als baumartig verschachtelte Tabelle dargestellten Bewertungsbaum top-down lesen und kennen diese Darstellung vermutlich bereits aus anderen Kontexten (R6). Beispielsweise werden Verzeichnisbäume häufig mit tabellarischen Darstellungen verknüpft. Aggregierte Ergebnisse, die zu inneren Knoten gehören, werden gemeinsam mit Einzelergebnissen der Blätter in einer neben dem Baum dargestellten Tabelle präsentiert. Solche Darstellungen können sowohl in webbasierten LMS als auch in Fat Clients durch in GUI-Bibliotheken vielfach verfügbare, sog. TreeTable-Widgets leicht realisiert werden (R2). Die

[5] Uns sind keine Autobewerter für Programmieraufgaben bekannt, die QTI unterstützen.
[6] Kursiv gesetzte Wortteile sind Elemente des Domänenmodells der Bewertungsvorschrift.

beiden *description*s enthalten auf Nutzeranforderung (bspw. als hover popup) anzeigbare Zusatzinformationen für Studierende bzw. Lehrkräfte.

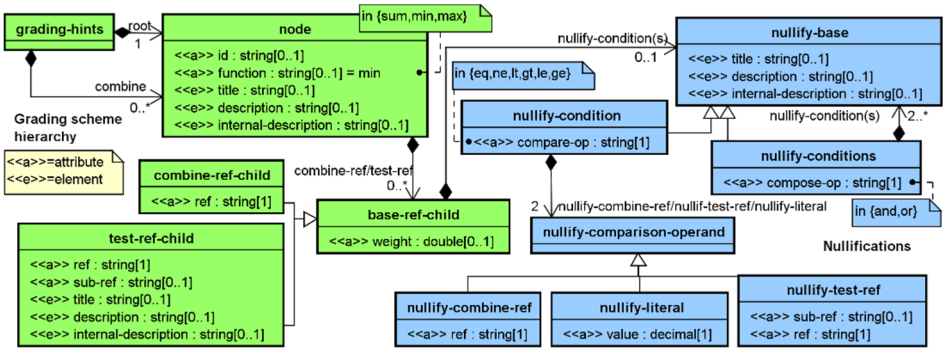



Abb. 2: UML-Domänenmodell des vorgeschlagenen XML-Formats. Links ist die Hierarchie ineinander verschachtelter Bewertungskategorien dargestellt, rechts die Elemente des Nullifikationen-Konzepts. Details des XML-Formats sind im ProFormA-Format-Repository[7] spezifiziert.

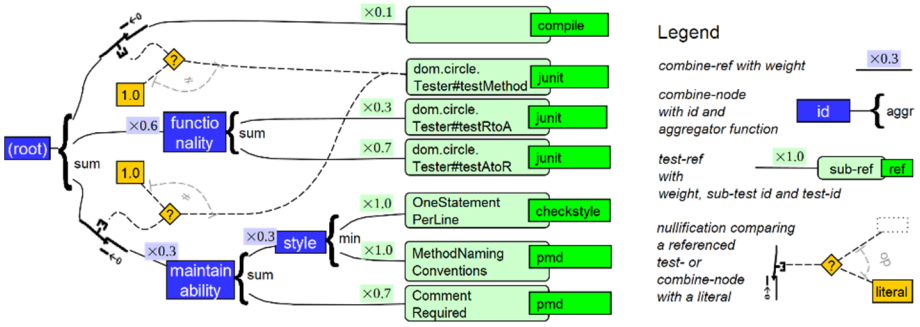



Abb. 3: Graphische Darstellung einer Bewertungsvorschrift im vorgeschlagenen Format. Die combine-ref-Kanten bilden einen Baum. Zusätzliche test-ref- und nullification-Verweise reichern den Baum zu einem Graphen an.

Abb. 3 zeigt das Bewertungsschema aus Tab. 1, wie es im vorgeschlagenen Format repräsentiert[8] wird. Testergebnisse sind rechts jeweils mit einer *ref*erence auf die Id des Tests (bspw. pmd) und der *sub-ref*erence auf die Id des Teiltests (bspw. CommentRequired) repräsentiert. Teiltests sind identifizierbare Teilprozesse eines Testwerkzeug-Durchlaufs. Ein Stilprüfer definiert üblicherweise die Namen einzelner Stilregeln als Teiltest-Ids; ein Unittest wird in einzelne Testfälle unterteilt, die z. B. im Falle von JUnit durch den Namen der Testmethode identifiziert werden. Der Compiler wird in unserem Fall nicht in Teiltests unterteilt. Die Testergebnisse werden an inneren, mit *id*s versehenen *combine*-Knoten über die ge*weight*eten Kanten aggregiert. Blattkanten referenzieren Tests und heißen deshalb *test-ref*s; Kanten zwischen inneren Knoten heißen *combine-*

[7] https://github.com/ProFormA/proformaxml
[8] Die entsprechende XML-Repräsentation findet sich in [Ga19].

*ref*s. Tab. 1 fordert Stilkonformität bzgl. Bezeichnern **und** der Anzahl der Anweisungen pro Zeile. Daher aggregiert der style-Knoten mit der min-Funktion.

Der in Tab. 1 erwähnte Sonderfall einer Trivialabgabe lässt sich durch sog. Nullifikationen abbilden, welche eine Ausprägung der Conditional-Aggregierung sind (vgl. Abschnitt 5). *Nullify-condition(s)* werden als Teil einer *test-ref*- oder *combine-ref*-Kante spezifiziert und bewirken, wenn die formulierte Bedingung erfüllt ist, die Auslöschung des von der Kante referenzierten Teilergebnisses, bevor es aggregiert wird. Ein *nullify-conditions*-Element definiert eine Auslöschungsbedingung, die aus mehreren *and*- oder *or*-verknüpften Teilbedingungen besteht. Einfache Vergleichsbedingungen besitzen im *nullify-conditon*-Element (beachten Sie das fehlende „s“ am Ende) einen *compare-op*erator, der per *nullify-combine-ref* oder *nullify-test-ref* referenzierte Teilergebnisse mit Zahlwerten (*nullify-literal*) vergleicht. Abb. 3 nutzt eine dedizierte JUnit-testMethod, um deren Ergebnis über gestrichelte Linien mit zwei *nullify-condition*s zu vernetzen, die das compile-Testergebnis bzw. das maintainability-Ergebnis dann auslöschen, wenn das Ergebnis von testMethod ≠ 1.0 ist. Die Java-Umsetzung von testMethod meldet einen Testerfolg (1.0), wenn keine Trivialabgabe vorliegt.

## 7 Evaluation

Zu den meisten der Anforderungen haben wir im obigen Text bereits argumentativ evaluiert, dass diese als erfüllt gelten können. Bzgl. R3 und R6 demonstriert Abb. 4, wie eine für Studierende nachvollziehbare LMS-Darstellung aussehen könnte, die zudem zuvor von einer Lehrkraft auf die Maximalpunktzahl 8 skaliert wurde. Die interaktiv navigierbare Darstellung hat der Student im Syntax- und im Style-Knoten „aufgeklappt“, um eine Erläuterung zur dort jeweils maximal erreichbaren Punktzahl von 0.80 bzw. 0.72 zu erhalten. Die Anordnung der Gewichte in Abb. 3 ist top-down-orientiert und damit günstig, wenn eine die Aufgabe wiederverwendende Lehrkraft leicht abwiechende Gewichtungen für Syntax, Funktion und Wartbarkeit definieren will (R3). Studierendenfreundlich (R6) wird es, wenn man die Gewichte wie in Abb. 4 automatisch zu Darstellungs- und Skalierungszwecken so zu den Blättern hin „propagiert“, dass die studentische Frage nach dem Einfluss der Einzelaspekte auf das Gesamtergebnis direkt ablesbar ist.

Die im vorliegenden Beitrag vorgeschlagene Erweiterung wurde und wird im Studienjahr 2018/19 in zwei konsekutiven Veranstaltungen zur Java-Programmierung mit je ca. 80 Studierenden im Bachelorstudiengang Angewandte Informatik der Hochschule Hannover real eingesetzt. Dabei wurden ca. 50 via Moodle zugängliche Graja[9]-Aufgaben auf das neue Format angepasst. Einige Aufgaben wurden dabei als randomisierbare Aufgaben umgesetzt, so dass hier bereits eine Übertragbarkeit der Ergebnisse auf neue Kontexte gezeigt werden konnte. Die Einsatzeffekte in Bezug auf Lernerfolge wurden nicht systematisch untersucht. Maßgebliches Ziel bei der Umstellung war, die Bewertung des Programmierstils fairer und damit lernförderlicher zu gestalten.

[9] http://graja.hs-hannover.de

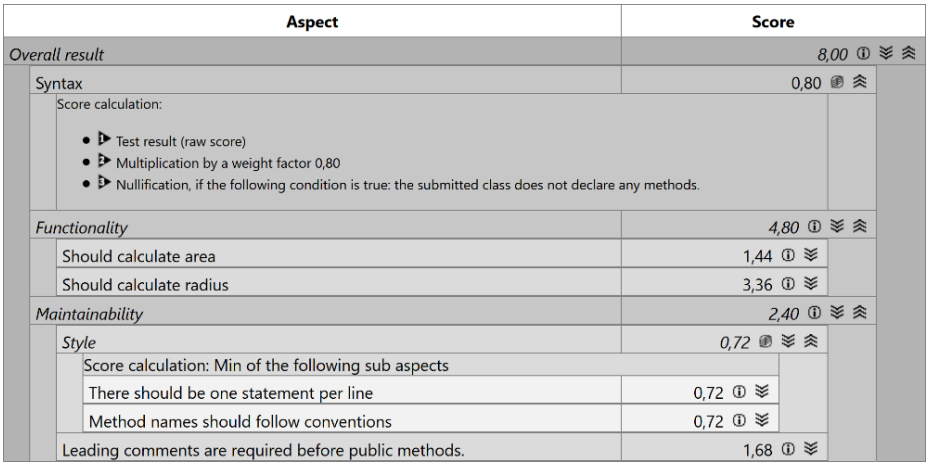



Abb. 4: Evaluation durch Realisierung einer Darstellungsmöglichkeit

Bzgl. des Forschungsziels der Übertragbarkeit des vorgeschlagenen Formats auf andere Bewertungssysteme und Programmiersprachen ist festzustellen, dass die Implementierungsarbeiten für weitere Systeme der o. g. Arbeitsgruppe noch nicht abgeschlossen sind. Die bei Durchstichen gewonnenen Erkenntnisse sind bereits in den Vorschlag eingeflossen, der inzwischen Teil der Version 2.0 des ProFormA-Formats ist und sich somit nahtlos in ein übergreifend nutzbares Aufgabenformat einfügt.

## 8 Ausblick

Will man auf der Grundlage unserer Ergebnisse nun ein gutes Bewertungsschema entwerfen, so kann man bspw. ein iteratives Forschungsdesign nutzen, wie es etwa in [St16] zur Entwicklung eines Bewertungsschemas für Codequalität beschrieben wird. Um zu einer bestimmten Bewertungsdimension wie der „Syntax“ in Tab. 1 eine umfassende Liste von Deskriptoren zu erhalten, die womöglich mit Exemplaren angereichert sind, kann man dem Beispiel von [JM15] folgen und viele ehemalige Klausuren durchforsten.

Grader und LMS müssen i. d. R. zur Unterstützung des vorgeschlagenen Bewertungsschemaformats angepasst werden. Indem wir die Eigenheiten vieler existierender Systeme berücksichtigt haben, gehen wir von überschaubaren Aufwänden zur Integration des neuen Formats in bestehende Systeme aus, welche idealerweise entlang des von der ProFormA-Arbeitsgruppe spezifizierten Formats für Bewertungsaufträge (submission) und des Antwortformats (response) erfolgt. Beide Formate[7] basieren auf dem hier vorgeschlagenen Bewertungsformat. Das submission-Format unterstützt Anforderung R3 dadurch, dass die in einer Aufgabe (task) festgelegte Bewertungsvorschrift durch eine eigene Vorschrift der Lehrkraft ersetzt werden kann.

Für die wertvollen Diskussions- und Entwicklungsbeiträge der Mitglieder der o. g. Arbeitsgruppe danke ich wie folgt namentlich und herzlich: Karin Borm, Peter Fricke, Elmar Ludwig, Oliver Müller, Uta Priss, Paul Reiser, Oliver Rod.

Literaturverzeichnis

[AB06] Attali, Y.; Burstein, J.: Automated Essay Scoring With e-rater® V.2. Journal of Technology, Learning, and Assessment, 4(3), 2006.

[Bi99] Biggs, J.: Enhancing teaching through constructive alignment. Higher Education, 32(3), 347-364, 1999.

[Bo17] Bott, O. J.; Fricke, P.; Priss, U.; Striewe, M.: Automatisierte Bewertung in der Programmierausbildung. Digitale Medien in der Hochschullehre, ELAN eV, Waxmann, 2017.

[Ga15] Garmann, R.; Heine, F.; Werner, P.: Grappa - die Spinne im Netz der Autobewerter und Lernmanagementsysteme. In: DeLFI 2015. LNI 247, GI, 169-181, 2015.

[Ga16] Garmann, R.; Fricke, P.; Bott, O. J.: Bewertungsaspekte und Tests in Java-Programmieraufgaben für Graja im ProFormA-Aufgabenformat. In: DeLFI 2016. LNI 262, GI, 215-220, 2016.

[Ga19] Garmann, R.: Ein Format für Bewertungsvorschriften in automatisiert bewertbaren Programmieraufgaben., Forschungsbericht, Hochschule Hannover, urn:nbn:de:bsz:960-opus4-13432, 2019.

[JM15] Jara, N.; Madrid, M. M.: Bewertungsschema für eine abgestufte Bewertung von Programmieraufgaben in E-Klausuren. In: DeLFI 2015. LNI 247, GI, 233-239, 2015.

[JS07] Jonsson, A.; Svingby, G.: The use of scoring rubrics: Reliability, validity and educational consequences. Educational research review, 2(2), 130-144, 2007.

[PJ13] Panadero, E.; Jonsson, A.: The use of scoring rubrics for formative assessment purposes revisited: A review. Educational Research Review, 9, 129-144, 2013.

[Sa89] Sadler, D. R.: Formative assessment and the design of instructional systems. Instructional science, 18(2), 119-144, 1989.

[St15] Strickroth, S.; Striewe, M.; Müller, O.; Priss, U.; Becker, S.; Rod, O.; Garmann, R.; Bott, O.; Pinkwart, N.: ProFormA: An XML-based exchange format for programming tasks. eleed e-learning & education, 11(1), 2015.

[St16] Stegeman, M.; Barendsen, E.; Smetsers, S.: Designing a rubric for feedback on code quality in programming courses. In: Koli Calling, ACM, 160-164, 2016.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 115

# Automatische Aufgabengenerierung über Linked Open Data am Beispiel der Archäologie

Michael Striewe[1]

**Abstract:** Automatische Aufgabengenerierung stellt fachspezifische Anforderungen an die verwendbaren Techniken. In manchen Fächern erscheint dabei die Nutzung von Linked Open Data vielversprechend. Dieses Paper untersucht am Beispiel der Archäologie die technischen und inhaltlichen Herausforderungen bei der Generierung von Multiple-Choice-Aufgaben. Die Untersuchung kommt zu dem Ergebnis, dass insbesondere die Vollständigkeit und inhaltliche Gewichtung der verfügbaren Daten ein Hindernis für die Realisierung darstellt.

**Keywords:** E-Assessment, Aufgabengenerierung, Linked Open Data, Semantic Web

## 1 Einleitung

Eine wiederkehrende Herausforderung in der Übungs- und Prüfungsvorbereitung ist die Erzeugung hinreichend vieler Aufgaben. Gerade im Übungsbetrieb ist es wünschenswert, ein Thema in voller Breite abdecken und viele Aufgaben auf unterschiedlichen Kompetenzniveaus anbieten zu können. Einen Lösungsansatz bieten hier Verfahren zur automatischen Aufgabenerzeugung [FR13, GLZ18], die in E-Assessments ggf. auch zur dynamischen Generierung unmittelbar beim Aufruf einer Aufgabe nutzbar sind. Im Fach Mathematik ist es bspw. gängig, Zahlen auszutauschen um beliebig viele Instanzen einer Aufgabe zu erzeugen [Ma18]. In textlastigen Fächern können Verständnisfragen zu Texten automatisch erzeugt werden [IGY13, MJ12]. Auf verschiedene Geistes- und Naturwissenschaften treffen beide Charakterisierungen jedoch nicht vollständig zu, wenn neben Zahlen, Formeln und Texten auch Abbildungen aller Art oder reale Gegenstände zum Einsatz kommen und Kompetenzen wie die sichere Beherrschung von Faktenwissen sowie die Anwendung von Kategorien und Konzepten geprüft werden sollen. Genau dies ist auch für die hier exemplarisch betrachtete Archäologie der Fall.

Die maschinenlesbare Abbildung von Fakten, Konzepten und Klassifikationen ist der Kern des „Semantic Web“, indem Informationen so mit Metadaten annotiert werden, dass sie verknüpft und ausgewertet werden können [SBH06]. Grundlage dafür sind Ontologien, die Entitäten und ihre möglichen Eigenschaften und Verknüpfungen beschreiben. Es existieren sowohl freie und thematisch offene derartige Datenbanken[2], als auch fachspezifische Datenbanken [La15]. Diese Datensammlungen mit Querverweisen untereinander werden auch als „Linked (Open) Data“ bezeichnet [Bi08].

[1] Universität Duisburg-Essen, paluno - The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, michael.striewe@paluno.uni-due.de
[2] Beispielsweise https://www.wikidata.org/ oder https://wiki.dbpedia.org/

doi:10.18420/delfi2019_104

Angesichts der freien Verfügbarkeit von Daten und standardisierten Technologien stellt sich die Frage, ob es möglich ist, sinnvolle Aufgaben für die Archäologie automatisiert zu erzeugen. Unbeschadet der theoretischen Machbarkeit wird die praktische Realisierbarkeit. aus technischer Sicht davon beeinflusst, ob Daten schnell genug verfügbar sind und aus fachlicher Sicht dadurch, ob die verfügbaren Daten korrekt sind, hinreichend vollständig um anspruchsvolle Aufgaben zu erstellen, und fachlich ausgewogen genug um ein Themengebiet in der nötigen Breite und Tiefe abzudecken. Letztlich steht hinter diesen Fragen die Evaluation der Nutzbarkeit von Linked Open Data Ressourcen zum Zwecke des E-Assessment. Da sich der Datenbestand in editierbaren Datenbanken naturgemäß jederzeit ändern kann, sind wertende Aussagen jedoch immer nur für einen gegebenen Zeitpunkt möglich. Alle Aussagen in diesem Paper beziehen sich daher auf einen Untersuchungszeitraum von Mitte Januar bis Mitte Februar 2019.

### 1.1 Verwandte Arbeiten

Arbeiten zur Aufgabengenerierung auf Basis von Linked (Open) Data bzw. Ontologien zeigen die grundsätzliche Umsetzbarkeit des Ansatzes [Al11, Sh19]. Die konkrete Nutzung offener, allgemeiner Datenbanken wurde ebenfalls bereits untersucht [Fo12, St18], ohne jedoch eine fachspezifische Evaluation durchzuführen. Allgemeine Ansätze, um die Qualität von Linked Open Data zu bewerten [Za16] oder systematisch zu verbessern [Ha18] gibt es ebenso wie Ansätze, die Qualität von offenen Daten in Relation zu einem bestimmten Einsatzzweck zu untersuchen [MZF17]. Untersuchungen mit Fokus auf die Aufgabengenerierung sind jedoch nicht verbreitet.

### 1.2 Technischer Hintergrund

Ein standardisiertes Format für Linked Data ist das „Resource Description Framework“ (RDF)[3]. Die semantische Beschreibung von Daten basiert auf Tripeln der Form „Subjekt – Prädikat – Objekt“, wobei Subjekt und Objekt Entitäten sind, während das Prädikat einer semantischen Beziehung in der zugrundeliegenden Ontologie entspricht. Das Subjekt könnte bspw. ein archäologisches Fundstück sein und das Objekt ein geographischer Ort. Ein Prädikat könnte dieser Beziehung dann die Bedeutung „Fundort“ zuweisen, ein anderes Prädikat dagegen die Bedeutung „Aufbewahrungsort“. Abfragen auf derartigen Daten können die ebenfalls standardisierte Abfragesprache SPARQL verwenden. Mit dieser können als Fortführung des o.g. Beispiels über eine Abfrage alle Objekte gefunden werden, für die derselbe Fundort oder Aufbewahrungsort hinterlegt ist. Über Mengenoperationen, Unterabfragen und Filter können auch komplexere Zusammenhänge zwischen Datenbeständen ausgewertet werden.

Die Umsetzung in diesem Paper basiert auf dem E-Assessment-System JACK, in dem Autoren Aufgaben mit Platzhaltern erstellen können, die dynamisch belegt werden. Die

[3] https://www.w3.org/RDF/

dazu benötigten Belegungsvorschriften können auch in Form von SPARQL-Abfragen angegeben werden, die beim Laden einer Aufgabe ausgeführt werden [St18].

## 2 Fallbeispiel 1: Historische Geographie

In der historischen Geographie lernen Studierende antike Städte, archäologische Stätten sowie historische Landschaftsbezeichnungen kennen und müssen sich dazu eine große Menge Faktenwissen aneignen, um bspw. im weiteren Studienverlauf in der provinzialrömischen Archäologie in der Lage zu sein, alle römischen Provinzen sowie wichtige Städte und bekannte Fundstätten in diesen Provinzen namentlich benennen und korrekt zuordnen zu können. Entsprechende Übungen sind daher Teil der Einführungsphase des Studiums. Eine einfache Umsetzung einer passenden Aufgabe benennt eine römische Provinz und fragte im Multiple-Choice-Verfahren nach der Zugehörigkeit von vier Städten. Von diesen kann eine oder mehrere korrekt sein.[4]

Zur Abfrage der benötigen Daten von Wikidata kann eine SPARQL-Anfrage (Abb. 1) verwendet werden, die im SELECT-Teil definiert, dass Paare von Orten und Provinzen gesucht werden. Im WHERE-Teil fragt sie zunächst Entitäten ab, die zur Kategorie Stadt (Q515) oder archäologische Stätte (Q839954) gehören. Ferner muss für diese Entitäten eine Zuordnung zu einem administrativen Gebiet vorliegt (P131) und bei diesem muss es sich um eine antike römische Provinz (Q182547) handeln. Filter begrenzen das Ergebnis auf Orte und Provinzen, für die die deutsche Bezeichnung vorliegt.

```
SELECT DISTINCT ?cityLabel ?provLabel WHERE {
  {?city wdt:P31 wd:Q515.} UNION {?city wdt:P31 wd:Q839954.}
  ?city wdt:P131 ?prov. ?prov wdt:P31 wd:Q182547.
  ?prov rdfs:label ?provL. ?city rdfs:label ?cityL.
  FILTER (langMatches( lang(?provL), "de" ) )
  FILTER (langMatches( lang(?cityL), "de" ) )
  SERVICE wikibase:label {bd:serviceParam wikibase:language "de".}
}
```

Abb. 1: SPARQL-Anfrage zur Abfrage von Städten und archäologischen Stätten in römischen Provinzen aus der Wikidata-Datenbank.

Aus dem Ergebnis wählt JACK bei Aufruf der Aufgabe zufällig vier Paare aus. Die Provinz aus dem ersten Paar wird im Aufgabentext angegeben. Die vier Orte aus allen Paaren bilden in zufälliger Reihenfolge die Antwortoptionen.

Die Laufzeit der Abfrage beträgt weniger als eine Sekunde und ist damit hinreichend schnell für die dynamische Aufgabengenerierung im Live-Betrieb. Alle erhaltenen Paare von Orten und Provinzen sind inhaltlich korrekt. Insgesamt liefert die Abfrage Paare mit 55 Orten in 16 Provinzen, was eine überraschend geringe Zahl darstellt, da tatsächlich deutlich mehr Orte und römische Provinzen in Wikidata erfasst sind. Grund ist, dass die Eigenschaft P131 in Verbindung mit einer römischen Provinz für die meisten Orte nicht

[4] Eine Beispielaufgabe ist verfügbar unter https://jack-demo.s3.uni-due.de/jack2/demo?id=63144

gesetzt ist. Die verfügbaren Daten sind demnach deutlich unvollständig. Kombinatorisch lassen sich aus den 55 Paaren gut 340.000 individuelle Aufgaben erzeugen. Der Wert dieser Zahl wird jedoch durch einen ungünstigen inhaltlichen Schwerpunkt gemindert: Von den 55 Paaren beziehen sich 37 auf die Provinz „Germania inferior“, so dass mehr als 2/3 der Aufgaben diese Provinz im Aufgabentext nennen werden. Die Breite des Themas wird so nicht angemessen reflektiert. Zufriedenstellend lösbar sind beide Probleme, indem fehlende Informationen zur Verknüpfung von Orten und Provinzen in Wikidata nachgetragen werden, um Umfang und Vollständigkeit der Daten zu erhöhen.

Eine andere Möglichkeit, Zuordnungen zwischen Orten und Provinzen zu ergänzen besteht darin, diese Information automatisch aus geographischen Daten abzuleiten. Koordinaten für Orte stehen dazu über verschiedene Quellen zur Verfügung. Vollständige und genaue antike Grenzverläufe sind jedoch in keiner allgemeinen oder fachspezifischen Quelle verfügbar, so dass dieser Ansatz nicht realisierbar ist.

## 3 Fallbeispiel 2: Bestimmungsübung

Neben Faktenwissen gehört die Bestimmung von Objekten hinsichtlich fachlicher Klassifikationen zum grundlegenden Handwerkszeug eines Archäologen. Übungen, in denen Studierende bspw. Keramikgefäße aufgrund ihrer Form und Bemalung in fachliche Kategorien einordnen müssen, gehören daher ebenfalls zur Grundausbildung in der Archäologie. Die dazu benötigten Daten sind etwas komplexer als im ersten Fallbeispiel, da die Beziehung zwischen einem Objekt und mehreren Klassifikationen bekannt sein muss und die zu bestimmenden Objekte zudem als Bild vorliegen müssen. Eine beispielhafte Umsetzung[5] einer passenden Aufgabe kann identisch zum ersten Fallbeispiel erfolgen: Es kommt wieder das Multiple-Choice-Verfahren zum Einsatz und es wird im Aufgabentext eine Klassifikation genannt, zu der eine oder mehrere der diesmal als Bild gegebenen Antwortoptionen passen. Die Aufgabe in einer Lernzieltaxonomie (z. B. [Kr02]) allerdings höher einzustufen als das erste Fallbeispiel, da Konzeptwissen angewandt werden muss, um die Aufgabe richtig zu lösen.

Da weder das benötigte Bildmaterial noch die fachlichen Klassifikationen über Wikidata zur Verfügung stehen, erfolgt die Realisierung über die Abfrage einer Fachdatenbank[6]. Die dabei verwendete SPARQL-Abfrage (Abb. 2) ist eine sog. Federated Query, die drei Datenquellen miteinander verknüpft. Im SELECT-Teil werden die gesuchten 4-Tupel (Objekt-ID, Bildreferenz, zwei Klassifikationen) eingeführt. Aus der Kerameikos-Datenbank werden dann im WHERE-Teil Objekte mitsamt der zugehörigen Abbildung (depiction) sowie ihre Klassifikation bezüglich Form (hasShape) und Bemalung (P32_used_general_technique) abgefragt. Die beiden letzten Informationen liegen in dieser Datenbank jedoch nur in Form einer Entitäten-URI vor, die über eine andere Datenbank in einen Namen aufgelöst werden muss. Dabei sind nicht alle Objekte in der

[5] Eine Beispielaufgabe ist verfügbar unter https://jack-demo.s3.uni-due.de/jack2/demo?id=68081
[6] http://www.kerameikos.org/

Kerameikos-Datenbank über Begriffe aus derselben Datenbank klassifiziert, weshalb der WHERE-Teil aus einer Vereinigung der Ergebnismenge aus zwei Unterabfragen besteht. Die erste Unterabfrage nutzt die Datenbank des British Museum zur Auflösung der Referenzen, die zweite die Datenbank des J. Paul Getty Trusts[7]. Die zusammengeführte Ausgabe wird abschließend zufällig sortiert und auf 20 Elemente begrenzt.

```
SELECT ?o ?p ?sl ?tl WHERE {
 { ?o kon:hasShape ?s. ?o foaf:depiction ?p. ?o crm:P32_used_general_technique ?t.
  SERVICE <https://collection.britishmuseum.org/sparql> {
   ?s skos:prefLabel ?sl.  ?t skos:prefLabel ?tl.
  } } UNION
 { ?o kon:hasShape ?s. ?o foaf:depiction ?p. ?o crm:P32_used_general_technique ?t.
  SERVICE <http://vocab.getty.edu/sparql> {
   ?s xl:prefLabel ?slf.  ?slf xl:literalForm ?sl.
   ?t xl:prefLabel ?tlf.  ?tlf xl:literalForm ?tl.
   FILTER (lang(?sl) = "en")  FILTER (lang(?tl) = "en")
  } } } ORDER BY RAND() LIMIT 20
```

Abb. 2: SPARQL-Anfrage zur Abfrage von antiken Keramikobjekten und ihrer Klassifikation aus der Kerameikos-Datenbank.

Die Laufzeit der Abfrage beträgt über 45 Sekunden, was die Live-Generierung von Aufgaben träge erscheinen lässt. Die abgefragten Daten sind fachlich korrekt, aber unvollständig: Von den ca. 320 Objekten ist für 80 kein Bild hinterlegt, so dass diese für die Aufgabe nicht verwendbar sind. Zwei der Bildverweise sind zudem fehlerhaft, so dass sporadisch lückenhafte Aufgaben erzeugt werden. Ebenfalls problematisch ist, dass die zwei verschiedenen Klassifikationen der beiden zusätzlichen Datenbanken nicht automatisch miteinander abgeglichen werden können. Antwortoptionen können daher von JACK nicht zuverlässig als korrekt erkannt werden. Derartige Fehler können ausgeschlossen werden, wenn nur die Objekte eines Museums verwendet werden. So wird jedoch der Vorteil der Verfügbarkeit vernetzter Daten faktisch aufgehoben.

## 4 Bewertung und Fazit

Die in diesem Paper durchgeführte Analyse kommt auf der praktischen Ebene zu einem negativen Fazit: Während die Qualität der verfügbaren Daten hinreichend gut ist, stehen in den untersuchten Fällen nicht ausreichend viele bzw. nicht hinreichend vollständige Daten zur Verfügung, um das jeweilige Thema über generierte Aufgaben sinnvoll abdecken zu können. Diese Aussage ist jedoch spezifisch für die gewählten Beispiele, da für andere fachliche Bereich vollständigere Daten zur Verfügung stehen könnten. Eine genaue Analyse für das jeweilige Thema ist daher unerlässlich. Trotzdem kann geschlossen werden, dass sich vor allem Umfang und Vollständigkeit der verfügbaren Linked Open Data Ressourcen noch verbessern müssen, bevor diese uneingeschränkt für die automatische Aufgabengenerierung in der Archäologie eingesetzt werden können.

[7] https://collection.britishmuseum.org/resource/sparql und http://vocab.getty.edu/sparql

## Literaturverzeichnis


[Al11] Al-Yahya, M.: OntoQue: A Question Generation Engine for Educational Assesment Based on Domain Ontologies. In: 11th IEEE International Conference on Advanced Learning Technologies, 393-395 (2011).

[Bi08] Bizer, C. et al.: Linked Data on the Web (LDOW2008). Proceedings of the 17th international conference on World Wide Web, 1265-1266, 2008.

[Fo12] Foulonneau, M.: Generating Educational Assessment Items from Linked Open Data: The Case of DBpedia. In: The Semantic Web: ESWC 2011 Workshops, 16-27, 2012.

[FR13] Foulonneau, M.; Ras, E.: Assessment Item Generation, the way forward. In: CAA 2013 International Conference, 2013.

[GLZ18] Gierl, M., Lai, H., Zhang, X.: Automatic Item Generation. In Encyclopedia of Information Science and Technology, 4th Edition (pp. 2369-2379). IGI Global, 2018.

[Ha18] Hadhiatma, A.: Improving data quality in the linked open data: a survey. In: Journal of Physics: Conference Series, 978, 2018.

[IGY13] Iwane, N., Gao, C., Yoshida, M.: Question Generation for Learner Centered Learning. In: 13th IEEE Int. Conference on Advanced Learning Technologies, 330-332, 2013.

[Kr02] Krathwohl, D.: A Revision of Bloom's Taxonomy: An Overview. In: Theory Into Practice, 41:4, 212-218, 2002.

[La15] Langner, M.: Archäologische Datenbanken als Virtuelle Museen. In: Digital Classics Online 1, 2015, 46-70, https://journals.ub.uni-heidelberg.de/index.php/dco/article/view/20314/14271

[MJ12] Mostow, J.; Jang, H.: Generating Diagnostic Multiple Choice Comprehension Cloze Questions. In: Proc. Workshop on Innovative Use of NLP for Building Educational Applications, 136–146, 2012.

[Ma18] Massing, T. et al.: E-Assessment Using Variable-Content Exercises in Mathematical Statistics. In: Journal of Statistics Education, 26(3), 174-189, 2018.

[MZF17] Mocnik, F.-B.; Zipf, A.; Fan, H.: Data Quality and Fitness for Purpose. In: 20th AGILE Conference on Geographic Information Science, 2017.

[SBH06] Shadbolt, N.; Berners-Lee, T.; Hall, W.: The Semantic Web Revisited. In: IEEE Intelligent Systems 21(3): 96-101, 2006.

[Sh19] Shi, W. et al.: A Framework for Automatically Generating Medical Quizzes with Multi-media Contents Based on Linked Data. In: Advances on Broadband and Wireless Computing, Communication and Applications, 147-158, 2019.

[St18] Striewe, M.: Dynamic Generation of Assessment Items Using Wikidata. In: Proc. of the 21st Int. Technology Enhanced Assessment Conference (TEA 2018), im Druck.

[Za16] Zaveri, A. et al.: Quality assessment for Linked Data: A Survey. In: Semantic Web, 7(1), 63-93, 2016.

# Digitalisiertes Feedback für verschiedene Frageformate von Online-Übungsaufgaben

Miriam Weigel[1], Katja Derr[1], Reinhold Hübl[1] und Tatyana Podgayetskaya[1]

**Abstract:** Im Rahmen des Hochschulverbundprojektes „optes – Optimierung der Selbststudiumsphase“ werden an der Dualen Hochschule Baden-Württemberg Mannheim Online-Mathematikkurse entwickelt. Eine besondere Rolle spielt der Einsatz digitalisierter Übungsaufgaben und die verschiedenen Rückmeldungsformate: Die eLearning Plattform ermöglicht sowohl eine automatisierte Auswertung von Online-Fragen als auch individuelles Feedback von eDozierenden. Der Artikel behandelt die Vor-und Nachteile verschiedener Fragetypen; Insbesondere wird der Fragetyp STACK vorgestellt, der ein qualitatives Feedback ermöglicht.

**Keywords:** formatives eAssessment, Online-Kurse, digitalisiertes Feedback, STACK-Fragetyp

## 1 Einleitung

Die heterogenen Vorkenntnisse von Studienanfänger/-innen stellen eine große Herausforderung für die Hochschullehre dar, insbesondere in den mathematikaffinen Grundvorlesungen. Um der Heterogenität entgegenzusteuern werden im Rahmen des Hochschulverbundprojektes „optes“ verschiedene Online-Angebote für die Studienvorbereitung und Studieneingangsphase entwickelt. Am Standort Mannheim gibt es ein Vorkursprogramm bestehend aus Online-Kursen, die im Selbststudium oder in Betreuten eLearning-Kursen bearbeitet werden. So soll es den angehenden Studierenden bereits vor Studienbeginn ermöglicht werden, Wissenslücken zu schließen und Defizite auszugleichen.

In diesem Beitrag wird das im Projekt verfolgte Konzept des automatisierten Feedbacks erläutert. Die technische Unterstützung durch die Lernplattform ILIAS ermöglicht verschiedene Fragetypen für die Aufgabenerstellung. Neben den Fragetypen Multiple Choice und Lückentext kommt oft der Fragetyp STACK zum Einsatz (vgl. [San13]). Diese Fragetypen ermöglichen ein qualitatives, automatisiertes Feedback, jedoch ist es nicht möglich Beweise oder Rechenwege vollumfänglich zu überprüfen. In den Betreuten eLearning-Kursen werden daher zusätzlich zu den Online-Tests PDF-Übungsblätter bereitgestellt. Die Studierenden laden ihre handschriftlichen Lösungen als PDF-Datei auf die Lernplattform hoch und die eDozierenden korrigieren diese online.

[1] Duale Hochschule Baden-Württemberg, Projekt optes, Coblitzallee 1-9, 68163 Mannheim

doi:10.18420/delfi2019_252

## 2 Unterschiedliche Fragetypen im Vergleich

In den Online-Kursen der DHBW Mannheim kommen die Fragetypen Multiple Choice, Lückentext und STACK am häufigsten zum Einsatz.

Multiple Choice-Aufgaben sind niederschwellig zu bedienen und verlangen von den Lernenden keine besonderen eLearning-Vorkenntnisse, Testteilnehmer/-innen können bei der Eingabe keine Flüchtigkeits- oder Schreibfehler machen. Die Eindeutigkeit der Eingabe führt zu größerer Fairness bei der Bewertung (vgl. [New02]). Auf der anderen Seite lässt die Einschränkung der Antwortmöglichkeiten auf eine überschaubare Auswahl keinen Raum für eigene Ansätze (vgl. [Mar99]). Es besteht die Gefahr, dass sich die Lernenden weniger auf das Wissen der richtigen Antwort als das Ausschließen der Distraktoren konzentrieren. Hinzu kommt, je nach Anzahl der Distraktoren, eine gewisse Ratewahrscheinlichkeit. Die Beantwortung vieler kurzer MC-Aufgaben kann auf Dauer eintönig sein und den Fokus auf die Leistungsmessung lenken, anstatt zum Nachdenken anzuregen. Auch wenn Multiple Choice-Fragen seit vielen Jahren in großen landesweiten und internationalen Leistungsmessungen zum Einsatz kommen, wurde immer wieder in Frage gestellt, ob überhaupt komplexere Gedankengänge, geschweige denn Problemlösestrategien getestet werden können (vgl. [Cra09]).

Im Gegensatz zu Multiple Choice-Fragen gibt es bei Lückentextfragen keine Antwortsuggestion. Bei den Lückentextfragen wird ein Begriff oder eine Zahl in ein Feld eingegeben. Die Bedienung ist ähnlich niederschwellig wie bei geschlossenen Formaten, ein Vorteil ist, dass die Rate-Wahrscheinlichkeit wegfällt. Ist die Lösung ein Wort, besteht allerdings die Gefahr, dass Testteilnehmer einen Begriff eingeben, der vom Autor nicht vorgesehen, aber dennoch eine korrekte Antwort auf die Frage ist. Diese Gefahr der Uneindeutigkeit ist bei numerischen Lücken nicht gegeben, Missverständnisse bei der Eingabe können aber dennoch entstehen. Darum ist zu berücksichtigen, dass jede potentiell richtige Antwort auch in der Aufgabenstellung hinterlegt ist (z.B. alternative Schreibweisen, Reihenfolgen von Begriffen, Kürzungsregeln etc.).

Hier verschafft das Plug-In STACK Abhilfe, das durch ein Computer Algebra System (CAS) an das Test-und Aufgabentool angebunden ist. In das Eingabefeld der STACK-Frage können Terme eingegeben werden, die anschließend mit Hilfe des CAS Maxima auf ihre mathematischen Eigenschaften geprüft und mit spezifischem Feedback versehen werden. Dadurch werden Fragestellungen mit beliebig vielen Lösungen ermöglicht und die Studierenden erhalten ein automatisiertes, qualitatives Feedback zu ihren Eingaben. Außerdem können STACK-Fragen mit Hilfe von Zufallsvariablen randomisiert werden, sodass mehrere Varianten einer Aufgabe entstehen. Aus Lernersicht ermöglicht STACK die Eingabe von Termen und Formeln via ASCIIMATH Syntax, verbunden mit einer direkten Vorschau (und der Möglichkeit zur Korrektur dieser Eingabe). Da die Werte einer Aufgabe immer wieder neu berechnet werden, sind STACK-Aufgaben ideal für den Einsatz im Übungsbereich (vgl. [San13]). Jedoch ist die Aufgabenerstellung und Testphase deutlich zeitintensiver als bei herkömmlichen Fragetypen.

Für die Entwicklung von Aufgaben für das Selbststudium sollte je nach Lernziel und Anspruchsniveau über den Aufgabentyp entschieden werden. Darüber hinaus erscheint eine Mischung der Aufgabentypen zur Auflockerung des Lernprozesses als sinnvoll.

## 3 Automatisiertes Feedback mit Hilfe des Fragetyps STACK

Der Schwerpunkt bei der Aufgabengestaltung für das Online-Selbststudium bildet ein aufschlussreiches, automatisiertes Feedback (siehe [Wei19]). Hierfür ist der Fragetyp STACK prädestiniert. Anhand der Beispielaufgabe „Ermittlung der Nullstellen einer quadratischen Funktion“ (vgl. Abb. 1) wird im Folgenden der technische Hintergrund einer STACK-Frage erläutert.

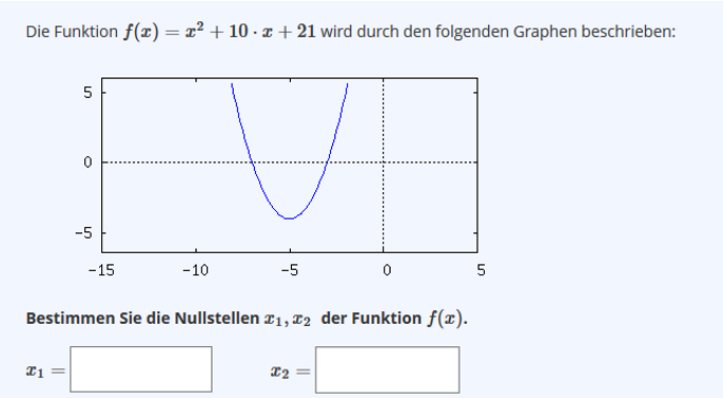



Abb. 1: Beispielaufgabe

Dem Fragenersteller wird unter den Einstellungen einer STACK-Frage ermöglicht, den Fragetext zu verfassen, Variablen zu definieren, einen Rückmeldebaum zu erstellen und individuelles Feedback einzugeben. Unter dem Feld Fragen-Variablen können CAS-Variablen definiert und verändert werden, die beispielsweise Zufallsvariablen ermöglichen. Diese Variablen sind in allen anderen Teilen der Aufgabe verfügbar. Speziell in dieser Aufgabe wurden zuerst die Nullstellen als Zufallsvariablen definiert. Anschließend wurden die Variablen p und q und die Funktion f(x) aus der Aufgabenstellung definiert. Je nachdem, welche zufälligen Werte für $x_1$ und $x_2$ gewählt werden, passt sich die Funktion f(x) in der Aufgabenstellung an. Mit tans1 und tans2 wurde jeweils die Musterlösung für die Eingabefelder benannt.

Zur Antwortüberprüfung wird ein Rückmeldebaum, der auf dem Binärbaumprinzip beruht, durchlaufen (vgl. Abb. 2). In den Knoten können mathematische Eigenschaften der studentischen Antwort überprüft werden. Die Verzweigungen ermöglichen die Eingabe von individuellem Feedback. Das Feld Feedback-Variablen erlaubt die studentische Eingabe zusammen mit den Aufgabenvariablen zu manipulieren, bevor der

Rückmeldebaum durchlaufen wird. Variablen, die hier definiert werden, können überall im Rückmeldebaum genutzt werden. In der Aufgabe aus Abb. 1 wurden die studentischen Eingaben (ans1 und ans2) als eine Menge zusammengefasst und anschließend die Schnittmenge aus dieser Menge und der Lösungsmenge gebildet. Danach wurde die Anzahl der Elemente der Schnittmenge ermittelt. Anschließend wurde aus den studentischen Eingaben eine quadratische Funktion g(x) mit Leitkoeffizienten 1 definiert (vgl. Abb. 2), die für das spezifische Feedback benötigt wird.

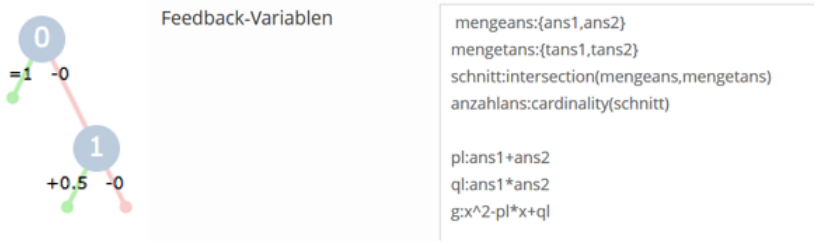



Abb. 2: Rückmeldebaum und Feedback-Variablen

Der Rückmeldebaum bildet das Zentrum einer STACK-Aufgabe. Es lassen sich beliebig viele Rückmeldebäume und Knoten zu einer Aufgabe erstellen. Jedoch sollte man aus Performance-Gründen und der Übersichtlichkeit wegen auf zu viele Verzweigungen und Rückmeldebäume verzichten. In der Beispielaufgabe können durch unsere zuvor definierten Feedback-Variablen Verzweigungen eingespart werden. Der erste Knoten in Abb. 2 überprüft, ob die Lösungsmenge und die Menge der studentischen Eingaben algebraisch übereinstimmen. Dadurch, dass jeweils Mengen gebildet und verglichen werden, ist die Reihenfolge der Eingabe der Lösungen irrelevant. Es ist daher nicht nötig zu überprüfen, ob die Nullstellen $x_1$ und $x_2$ vertauscht wurden. Sind die Lösungsmenge und die Eingabemenge gleich, so erhalten die Studierenden volle Punktzahl und das Feedback „Ihre Antwort ist korrekt“ mit einer Erläuterung zur Musterlösung. Stimmen Lösungsmenge und Eingabemenge nicht überein, so wird der zweite Knoten durchlaufen. Hier wird überprüft, ob die Anzahl der Elemente in der Schnittmenge 1 ist. Ist dies der Fall, so haben die Studierenden eine Nullstelle richtig eingegeben (die Reihenfolge der Eingabe von $x_1$ und $x_2$ spielt erneut keine Rolle) und es werden Teilpunkte vergeben. Die Studierenden erhalten das Feedback „Ihre Antwort ist teilweise korrekt. Sie haben eine Nullstelle richtig berechnet.“ zusammen mit einer Grafik, die die Musterlösung f(x) und die studentische Funktion g(x), die in den Feedback-Variablen definiert wurde, zeigt (vgl. Abb. 3). Durch die visuelle Aufbereitung der studentischen Eingabe können die Studierenden auf einen Blick erkennen, dass ihre eingegebenen Nullstellen nicht mit den geforderten Nullstellen übereinstimmen.

Beträgt die Schnittmenge im zweiten Knoten nicht 1, so wurde keine Nullstelle richtig berechnet und die Studierenden erhalten ein entsprechendes automatisiertes Feedback mit einem angepassten Schaubild. Zusätzlich wird bei jeder Antwortauswertung das allgemeine Feedback mit einer Erläuterung zur Musterlösung zur Verfügung gestellt.

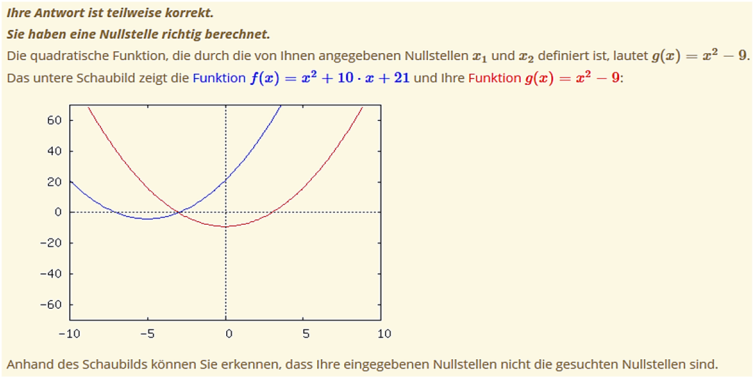


Abb. 3: Musterlösung und Feedback

## 4 Erste Erfahrungen

Seit 2014 werden an der DHBW Mannheim betreute eLearning-Kurse angeboten. Im Sommer 2018 wurde der Einsatz des Fragetyps STACK mit einer kleinen Gruppe von Studienanfänger/-innen pilotiert (n = 19), die an fünf Präsenzterminen an der DHBW Mannheim teilnahmen. Aufgrund der unbegrenzten Eingabemöglichkeiten war im Vorfeld eine intensive Testphase der STACK-Aufgaben unerlässlich um das gewünschte Frageverhalten zu erzielen. Beispielsweise reicht die Antwortüberprüfung „algebraisch äquivalent“ nicht aus, wenn ein Bruch gekürzt werden soll, da das bloße Abschreiben des ursprünglichen Bruchs volle Punktzahl erzielen würde.

In den Präsenz-Veranstaltungen bearbeiteten die Pilot-Teilnehmer Online-Tests mit STACK-Fragen. Die anwesenden Projektmitarbeiter/-innen konnten so direkt beobachten, wie die Lernenden mit den Aufgaben interagieren und ggf. direkt auf Probleme reagieren. Im Anschluss schilderten die Lernenden in kurzen Feedbackrunden ihre Eindrücke zum Umgang mit dem Fragetyp. Diese erste formative Evaluation zeigte eine gute Akzeptanz bei den Lernenden. Als besonders positiv wurde genannt, dass das Feedback eine detaillierte Musterlösung enthielt und dass jeder Testaufruf andere Aufgabenwerte erzeugt und sich auf diese Weise der Übungseffekt verstärkt. Allerdings hatten einige Vorkursteilnehmer/-innen anfangs Schwierigkeiten mit der Eingabesyntax. Eine kurze Anleitung und Testphase mit dem ASCII Format ist darum unerlässlich. Aus Sicht der Lehrenden wurde das Erlernen der Eingabesyntax sogar als positiver Nebeneffekt bewertet, da die Studierenden sich so auf den Umgang mit Mathematiksoftware vorbereiten können.

Nach diesen positiven Erfahrungen ist im Sommer 2019 der umfassende Einsatz von STACK mit allen Vorkursteilnehmern der Fakultät Technik geplant.

## 5 Zusammenfassung

Digitalisiertes Feedback ist zeit- und ortsunabhängig und daher von großem Nutzen für Lernende und Lehrende. Bei der Entscheidung, welcher Aufgabentyp zum Einsatz kommt, sind neben dem Lernziel, sowie dem Komplexitäts-und Schwierigkeitsgrad der Aufgabenstellung auch die meta-kognitiven Anforderungen zu berücksichtigen, die durch das technische Handling entstehen können. Die eher unflexiblen Multiple Choice-Aufgaben sind aufgrund ihrer Niederschwelligkeit bei der Eingabe für unerfahrene eLearning-Teilnehmer/-innen von Vorteil; jedoch birgt das Aufgabenformat ein hohes „Raterisiko“. Abhilfe schaffen Aufgabenformate, bei denen ein mathematischer Ausdruck eingegeben werden muss (z.B. Lückentext- oder STACK-Aufgaben). Aufgaben, die die Abgabe ganzer Rechenschritte oder mathematischer Lösungsansätze verlangen, erfordern die individuelle Korrektur und Rückmeldung einer Lehrperson, und sind daher nur im Rahmen betreuter Online-Vorkurse anwendbar.

Besonders wichtig für die Feedbackgestaltung von Online-Fragen ist es die Ziele im Lernprozess zu definieren: steht die Selbstkontrolle im Mittelpunkt, so reichen Frageformate mit einfachem und flachem Feedback aus. Wird ein qualitatives Feedback mit Fehleranalyse benötigt, dann bietet der Fragetyp STACK hervorragende Möglichkeiten für die Gestaltung von automatisierten und adaptiven Feedback.

Literaturverzeichnis

[Cos14] cosh cooperation schule:hochschule. Mindestanforderungskatalog Mathematik (2.0) der Hochschulen Baden-Württembergs für ein Studium von WiMINT-Fächern. Stand 2014. http://www.mathematik-schule-hochschule.de/images/Aktuelles/pdf/MAKatalog_2_0.pdf

[Cra09] Craven, P.: History and Challenges of e-assessment. The 'Cambridge Approach' perspective-e-assessment research and development 1989 to 2009. http://www.cambridgeassessment.org.uk.

[Mar99] Martinez, M. E.: Cognition and the Question of Test Item Format. Educational Psychologist, 34 (4), 1999.

[New02] Newstead, S.: Examining the examiners: Why are we so bad at assessing students? Psychology Learning and Teaching, 2002.

[San13] Sangwin, C.: Computer Aided Assessment of Mathematics Using STACK. Oxford University Press, Oxford, 2013.

[Wei19] Weigel, M. et al.: STACK-Aufgaben im formativen eAssessment: Einsatzmöglichkeiten des Feedbacks, 2019. Zenodo. http://doi.org/10.5281/zenodo.2563656.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 127

# Entwicklung eines modernen Lehr- und Lernkonzepts für Maschinelles Lernen mit Infrastructure-as-a-Service und Cloud-Computing

Patrick Palsbröker, Matthias Lutz, Carsten Gips und Matthias König[1]

**Abstract:** Im Rahmen des BMBF-geförderten Projektes „Machine Learning Campus Minden“ wurde ein Lehr- und Lernkonzept entwickelt, um den Schwerpunkt Maschinelles Lernen (ML) im Curriculum des Masterstudiengangs Informatik der Fachhochschule Bielefeld auszubauen. Dazu wird das Grundlagenmodul “Methoden des Maschinellen Lernens” zur Vermittlung des theoretischen Basiswissens angeboten. Darauf aufbauend können die Studierenden ihre Kenntnisse in einer Reihe von anwendungsorientierten Vertiefungsmodulen wie beispielsweise “Natural Language Processing” (NLP) oder “Data Science“ (DS) an Fragestellungen aus dem Forschungs- und Industriealltag erproben und erweitern. Um eine praxisnahe Ausbildung auf dem aktuellen technologischen Stand zu gewährleisten, bieten alle neuen Module die praktische Anwendung mittels modernen Cloud-Plattformen an. In diesem Beitrag wird das so entstandene Lehr- und Lernkonzept vorgestellt. Außerdem stellen wir die gesammelten Erfahrungen bei der Integration von Infrastructure-as-a-Service-Diensten in der Lehre dar und gehen sowohl auf die fachlichen als auch organisatorischen Anforderungen und Fallstricke ein sowie auf die Evaluation der eingesetzten Cloud-Anbieter in Hinblick auf die domänenspezifischen und praxisrelevanten Probleme.

**Keywords:** Lehrkonzept, Lernkonzept, Maschinelles Lernen, Cloud-Computing, Google Cloud Platform, Amazon Web Services, Microsoft Azure.

## 1 Einleitung

Maschinelles Lernen in seinen vielen Facetten erhält eine immer stärkere Bedeutung, weshalb es uns notwendig erschien, die Lehre in diesem Bereich deutlich auszubauen. Die Verarbeitung von großen Datenmengen erfordert sehr leistungsfähige Hardware, deren Anschaffung und Wartung gerade für Hochschulen eine Herausforderung darstellt [YF15], zumal die für das Maschinelle Lernen benötigte Hardware schnell veraltet. Nach Umfragen aus dem Jahr 2014 stellen lediglich ein Drittel der Hochschulen Rechencluster bereit, so dass es in Lehrveranstaltungen bei einer theoretischen Betrachtung der Methoden bleibt [ST14]. Wir stellen daher in diesem Artikel ein Lehr- und Lernkonzept zur praxisnahen Ausbildung im Bereich Maschinelles Lernen sowie ein Konzept zur flexiblen Integration von Cloud-Diensten im Rahmen der Lehre vor.

[1] FH Bielefeld Campus Minden, Artilleriestraße 9, 32427 Minden {Vorname.Nachname}@fh-bielefeld.de

doi:10.18420/delfi2019_274

## 2 Lehr- und Lernkonzept

Das im Rahmen des vom BMBF geförderten Projektvorhabens entwickelte Lehr- und Lernkonzept setzt sich, wie in Abb. 1 dargestellt, aus drei aufeinander aufbauenden Ebenen zusammen: Basismodul, Vertiefungsmodule und abschließende Masterarbeit.

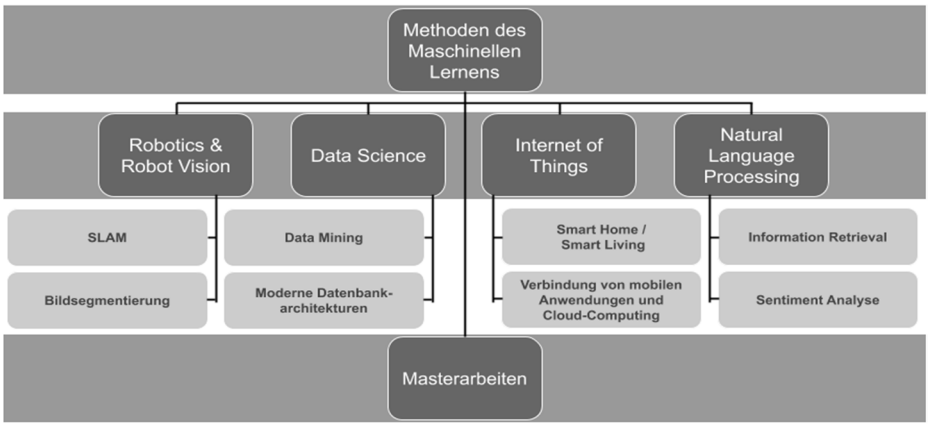



Abb. 1: ML-Camp Lehr- und Lernkonzept

Im Basismodul werden aktuelle Algorithmen und Methoden aus dem Bereich ML umfangreich vermittelt, wobei der Fokus auf Algorithmen des überwachten und unüberwachten Lernens liegt. Ein wichtiger Bestandteil sind speziell konzipierte Praktika, zu deren Bearbeitung größtenteils Cloud-Dienste genutzt werden. Das Grundwissen für den Umgang mit den eingesetzten Cloud-Plattformen wird in einem initial durchgeführten Workshop zum Semesterstart vermittelt. Da es bisher an Sach- und Lehrbüchern auf dem Markt zur Ausarbeitung von Vorlesungen mit Cloud-Nutzung mangelt [ST14], wurde auf verschiedene andere Quellen zurückgegriffen. Das Basismodul wird in jedem Semester als Wahlfach angeboten, um allen Studierenden die Möglichkeit zu geben, dieses in Vorbereitung auf die Vertiefungsmodule zu besuchen. Während in der Vorlesung primär Konzepte und der mathematische Hintergrund der Verfahren thematisiert werden, untersuchen die Studierenden in einem kleinen Projekt im praktischen Teil des Moduls eine Fragestellung aus dem Forschungs- oder Industriealltag und beschäftigen sich mit der Anwendung der ML-Algorithmen. In den anschließenden Semestern kann das erworbene Wissen in verschiedenen anwendungsorientierten Themenschwerpunkten vertieft werden. Basierend auf der aktuellen Forschungstätigkeit im Studiengang wurden hier zunächst Vertiefungsmodule mit Schwerpunkt auf Robotics & Robot Vision, Data Science, Internet of Things und Natural Language Processing & Information Retrieval entwickelt. In diesen Vertiefungsmodulen steht die Untersuchung von praxisnahen Fragestellungen im Rahmen von Projekten im Vordergrund. Angestrebt sind also die Vermittlung methodischen Vorgehens zur effizienten Problemlösung und Optimierung angewandter Verfahren wie beispielsweise aktuell in einem Studierendenprojekt, Generative Adversarial Networks

zur Generierung von Trainingsdaten für die Klassifizierung von gefälschten Unterschriften. Den Studierenden ist es in den Vertiefungsmodulen freigestellt, in welcher Form sie die verschiedenen Cloud-Dienste in ihre Projekte einbinden. Darauf aufbauend kann eine Masterthesis in diesem Bereich erstellt werden, beispielsweise in Kooperation mit einem Firmenpartner oder im Kontext passender Forschungsprojekte am Campus Minden.

## 3 Cloud Konzept

Die Module aus dem Lehrkonzept haben potenziell sehr unterschiedliche Anforderungen hinsichtlich der nötigen Vorkenntnisse, der Auswahl der ML-Algorithmen und der gewünschten Abstraktionsebene. Wie schon beschrieben, liegt im Basismodul "Methoden des maschinellen Lernens" ein starker Fokus auf den Funktionsweisen und dem mathematischen Hintergrund der behandelten Methoden. Aus diesem Grund wird hier in der Regel auf den Einsatz etablierter Bibliotheken verzichtet und eine eigene Implementierung durch die Studierenden erwartet, um einen tiefen Einblick in die verschiedenen Algorithmen zu erlangen. Auf der anderen Seite stehen Module wie NLP, die stark anwendungsorientiert sind und deren Anforderungen eher auf der Anwendung, Evaluation und Kombination diverser Best Practices sowie (Hyper-) Parameter-Tuning von ML-Algorithmen liegen. Aus diesem Grund wurde die Entscheidung getroffen, mindestens zwei Cloud-Plattformen einzubeziehen, die sich im Service-Modell unterscheiden: Während das Machine Learning Studio (MLS) von Microsoft als Software-as-a-Service zu interpretieren ist, entspricht das Amazon Web Services-Angebot einem IaaS-Dienst. Die Integration von Cloud-Computing im beschriebenen Kontext lässt sich noch als Novum in der Lehre ansehen, daher existieren wenig bis keine Richtlinien für die effektive Nutzung und keine Orientierung bei der Erstellung eines Konzepts. Zwar wurde Ende 2018 eine Richtlinie für Cloud-Computing von einer internationalen Expertengruppe veröffentlicht [Fo18], jedoch muss diese noch sukzessive Einhalt in die Curricula finden und sich beweisen. Viele angesehene Hochschulen führen zwar das Schlagwort Cloud in der Prüfungsordnung, nutzen diese jedoch als Tool zu 80% nicht in der Praxis [ST14]. Ein Grund für die träge Zuwendung zu Cloud-Diensten in Unternehmen wie auch in Hochschulen ist das Mietmodell samt Nebeneffekten: Organisation der Datensicherung, -übernahme und -rückgabe sowie Lokalisierung zum Zeitpunkt der Verarbeitung, und die Stabilität der Mietkosten in absehbarer Zeit und Intervenierbarkeit seitens des Personals. Dabei gilt es insbesondere das Prinzip der Transparenz zu gewährleisten. Zur Anonymisierung sensibler Daten lassen sich beispielsweise Hashfunktionen einsetzen und bezüglich der Datensicherheit ein zentrales Schlüsselverwaltungssystem etablieren. Laut einer Studie von Bitkom Research [PV17] verbessern sich aber auch diverse Faktoren durch Einführung eines durchdachten Cloud-Computing-Konzeptes, darunter fallen die Aspekte Anwendungssicherheit, Implementierungszeiten und der IT-Administrationsaufwand. Entgegen der allgemeinen Meinung, dass Daten in der Cloud als unsicher gelten, zeigt die Studie deutlich, dass die Daten in der Public Cloud gegenwärtig als sicher beurteilt werden. Personenbezogene Daten sind jedoch zu vermeiden, da

dadurch wegen der Datenschutzgrundverordnung ein größerer bürokratischer Aufwand festzustellen ist. Die Informationen zum Empfänger, Zweck der Datenverarbeitung und die Aufbewahrung der Daten sind dokumentationspflichtig. Aufgrund der Tatsache, dass führende Cloud-Anbieter in den USA lokalisiert sind, ist besondere Vorsicht zu wahren. Unter diesen Voraussetzungen ist es schwer, direkte Angebote einzuholen. Durch das EU-US Privacy Shield-Datenschutzübereinkommen und die Partnerschaft mit Resellern, welche sich auf Öffentliche Einrichtungen und deren Ansprüche bei Support und Abwicklung spezialisiert haben, wird dieser Prozess jedoch deutlich vereinfacht. Die Verantwortung obliegt jedoch weiterhin der Hochschule, daher werden in Absprache mit den Resellern ausschließlich Prozesse implementiert und Dienste verwendet, welche die Bedingungen der DSGVO einhalten.

# 4 Evaluation

## 4.1 Methodik und Betrachtung der Ergebnisse

Zur Evaluation des Lehr- und Lernkonzepts wurde am Ende jedes Moduls eine Umfrage mit allen teilnehmenden Studierenden durchgeführt. Die Ergebnisse sind in Abb. 2 dargestellt. Den Umfrageergebnissen kann man entnehmen, dass die Studierenden mit dem inhaltlichen Aufbau der Module gut zurechtkamen.

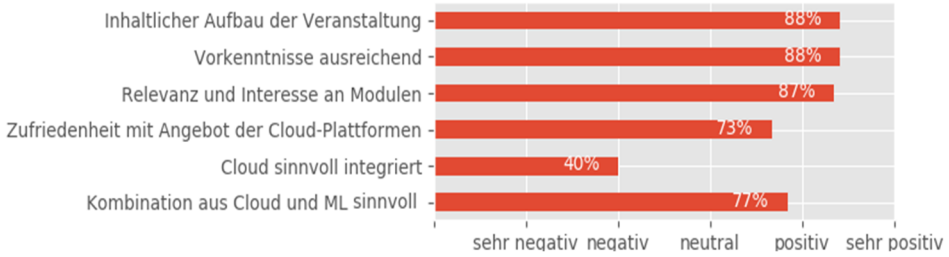



Abb. 2: Umfrageergebnisse

Es ist gelungen, das richtige Einstiegsniveau zu finden (88% der Teilnehmer gaben an, dass die Vorkenntnisse ausreichend waren zum Besuch der Module). Zudem zeigt sich, dass das Interesse an den Modulen wie erwartet sehr hoch ist und die vermittelten Inhalte für die Teilnehmer hohe Relevanz aufweisen. Den im Projekt verfolgten Ansatz, Rechenleistung flexibel über Cloud-Dienste anzumieten statt kosten- und wartungsintensive Hardware anzuschaffen, kann die überwiegende Mehrheit der Teilnehmer nachvollziehen (77%). Die Integration der Cloud-Dienste im Basismodul muss jedoch deutlich verbessert werden, lediglich 40% der Studierenden fanden den Einsatz der Cloud-Dienste im ersten Durchlauf gelungen. In Reaktion auf diese Umfrageergebnisse wurde bei der Wiederholung des Basismoduls eine hohe Priorität auf die stärkere und frühere Integration der Cloud-Dienste innerhalb des Semesters gelegt.

Beispielsweise fand eine Überarbeitung der Praktika statt, um mehr Aspekte der verschiedenen Dienste abzubilden. Außerdem findet zu Semesterstart nun ein Workshop als Pflichtveranstaltung statt, um den Studierenden den Zugang zu den verwendeten Cloud-Diensten zu erleichtern. Ob diese Änderungen sich positiv auswirken, wird sich bei den Umfragen nach dem aktuell laufenden Semester zeigen.

## 4.2 Evaluation der durch die Cloud-Plattformen entstandenen Kosten

Im Folgenden sollen die Kosten der eingesetzten Cloud-Dienste betrachtet werden, wobei zu beachten ist, dass pro Dienstleister eine maximale Anzahl von ca. 25 Studierenden das Angebot genutzt hat. Die Kosten der Cloud-Nutzung blieben dabei unter dem vorab kalkulierten Bedarf, da vor allem in der vorlesungsfreien Zeit wenig Nachfrage bestand. Für das Sommersemester 2018 sind so Kosten in Höhe von 2537,17 entstanden. Obwohl die hier betrachteten Dienste ähnliche Gesamtkosten erzeugt haben, setzen sich diese unterschiedlich zusammen. So fallen bei Microsoft Grundgebühren für jeden virtuellen Arbeitsplatz an. Dafür sind die Gebrauchskosten jedoch so gering, dass diese kaum ins Gewicht fallen. Bei AWS ist das Gegenteilige der Fall (siehe Abb. 3).

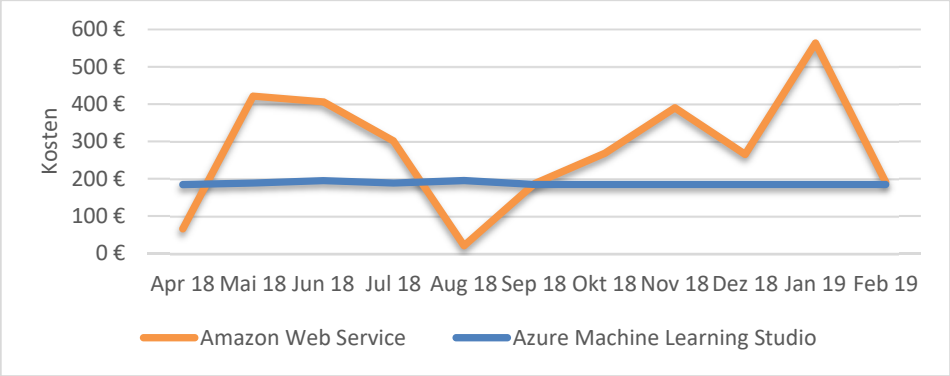



Abb. 3: Vergleich der Kosten pro Cloud-Dienst

Vergleicht man diese Kosten mit selbst angeschaffter Hardware, fallen diese sehr gering aus. Dies liegt hauptsächlich daran, dass im Lehreinsatz häufig stoßweise große Kapazitäten benötigt werden, beispielsweise während einer Lehrveranstaltung, es aber auch lange Zeiträume gibt, in der die Leistung nicht gebraucht wird. Die von uns überschlagenen Kosten, welche auf eingeholten Angeboten basieren, belaufen sich für einen Rechencluster mit vergleichbarer Leistung wie die von uns eingesetzten AWS Maschinen, bei 25 Teilnehmern auf ca. 3000 Euro im Monat. Dabei sind relevante Kostenfaktoren wie Strom und Wartung der Cluster noch nicht eingeschlossen, die jedoch auch einen erheblichen Anteil der Gesamtkosten ausmachen [YF15].

## 5 Zusammenfassung und Ausblick

Das in diesem Bericht beschriebene Konzept zur Ausbildung im Bereich Maschinelles Lernen wurde erfolgreich umgesetzt. Das Basismodul und verschiedene Vertiefungsmodule wurden erstmalig durchgeführt und sind auf große Resonanz unter den Studierenden des Masterstudiengangs Informatik am Campus Minden gestoßen. Im Projekt konnten wertvolle Erfahrungen im Einsatz von Cloud-Diensten für Maschinelles Lernen in der Lehre gesammelt werden, die in die Verbesserung der Module einfließen werden. Die daraus resultierenden Vorteile gegenüber der Anschaffung eigener Hardware, wie Wartungsfreiheit und verfügbare Ressourcen, stachen insgesamt hervor. Durch die Einführung von Cloud-Diensten kann die Qualität und Kosteneffektivität von Maschinellem Lernen im Lehreinsatz zunehmen, sowie der Energieverbrauch durch die dezentrale Virtualisierung abnehmen. Bisher wurden die Plattformen AWS und Azure eingesetzt. Im aktuell laufenden Semester wird das Angebot um die GCP ergänzt. Ein wichtiger Aspekt ist dabei die Kostenkontrolle und -vorhersage. Im bisherigen Verlauf dieses Projekts sind die Kosten unter den Prognosen geblieben. Speziell bei der Nutzung der GCP erscheint es sinnvoll zu evaluieren, inwieweit isolierte Lernumgebungen zum kostenlimitierten Experimentieren durch Cloud-Sandboxes eingesetzt werden können. Alternativ würde ein eigenes Management der Ressourcen in Verbindung mit einer restriktiven Konfiguration des Cloud Identity & Access Management eine sichere und isolierte Umgebung für die Cloud-Dienste darstellen.

## Danksagung

Die vorliegende Arbeit wurde finanziell im Rahmen des BMBF-Vorhabens "ML Camp - Machine Learning Campus Minden" unter dem Kennzeichen 01 IS 17081 gefördert.

Literaturverzeichnis

[Fo18] Foster, Derek, et al.: Cloud computing: developing contemporary computer science curriculum for a cloud-first future. Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education. ACM, 2018.

[YF15] Yaghmaei, Omid; Binesh, Fatemeh: Impact of applying cloud computing on universities expenses. IOSR Journal of Business and Management Volume 17.2 S. 42-47, 2015.

[PV17] Pols, Axel; Vogel, Marko: Bitkom Research GmbH, https://www.bitkom.org/sites/default/files/pdf/Presse/Anhaenge-an-PIs/2017/03-Maerz/Bitkom-KPMG-Charts-PK-Cloud-Monitor-14032017.pdf, Stand: 05.03.2019.

[ST14] Scherzinger, Stefanie; Thor, Andreas: Cloud-Technologien in der Hochschullehre – Pflicht oder Kür?. Datenbank-Spektrum 10/14 S. 131-134, 2014.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 133

# Einsatz von Virtual Reality zum Aufbau von Klassenmanagement-Kompetenzen im Lehramtsstudium

Axel Wiepke[1], Eric Richter[2], Raphael Zender[1] und Dirk Richter[2]

**Abstract:** Virtuelle Realitäten ermöglichen Trainingsszenarien, die in der realen Welt nur mit hohem Aufwand oder mit ethischen Bedenken durchgeführt werden können. Dazu gehört das Verhaltenstraining angehender LehrerInnen vor Schulklassen. Dieser Beitrag beschreibt eine für diesen Zweck entwickelte VR-Klassenraumsimulation, in der gezielt Störungen bei simulierten SchülerInnen ausgelöst werden können. Der Einsatz dieser Lösung für das realitätsnahe Training von Klassenmanagementkompetenzen bei Lehramtsstudierenden wird beschrieben und evaluiert, wobei ein Fokus auf dem angemessenen Umgang mit Unterrichtsstörungen liegt.

**Keywords:** Virtual Reality, Lehrerbildung, Simulation

## 1 Motivation

Virtuelle Realitäten (VR) sind durch die dreidimensionale Darstellung, ihren immersiven Charakter und die Möglichkeit einer direkten Interaktion für den Einsatz in verschiedenen Szenarien geeignet. So kann VR beispielsweise effektiv in Lehr- und Lernszenarien eingesetzt werden [Ze18]. Das Medium ermöglicht die Vermittlung von Handlungskompetenzen in emotional anspruchsvollen Situationen, die in der physischen Realität nur unzureichend, mit erheblichem Aufwand oder unter ethisch bedenklichen Bedingungen zu Trainingszwecken erzeugt werden können – beispielsweise zum Verhaltenstraining für angehende LehrerInnen in authentischen Klassenzimmern. In der Vergangenheit kamen zur Vorbereitung von Lehramtsstudierenden auf Störsituationen im Unterricht vor allem textlich beschriebene Situationen und Videos von Unterrichtsmitschnitten [PSE15] zum Einsatz. Insbesondere das erforderliche hohe Maß an Abstraktionsfähigkeit (z. B. in textuellen Situationen) sowie die abschwächende Wirkung der Distanz zum Geschehen (z. B. durch Unterrichtsvideos) erschweren jedoch den Transfer des Gelernten auf späteren realen Unterricht.

Diese Probleme adressiert das in Zusammenarbeit zwischen Informatik und Bildungsforschung entstandene VR-Klassenzimmer, das künftige LehrerInnen simulativ, praxisnah und zielgerichtet auf das Verhalten in Störsituationen vorbereitet [WP18].

[1] Universität Potsdam, Institut für Informatik und Computational Science, August-Bebel-Str. 89, 14482, Potsdam, vorname.nachname@uni-potsdam.de
[2] Universität Potsdam, Erziehungswissenschaftliche Bildungsforschung, Karl-Liebknecht-Straße 24-25, 14476, Potsdam, vorname.nachname@uni-potsdam.de

doi:10.18420/delfi2019_319

In diesem Beitrag werden das VR-Klassenzimmer sowie dessen Einsatz und Evaluierung im Lehramtsstudium an der Universität Potsdam vorgestellt und diskutiert.

## 2 Verwandte Arbeiten

Durch die steigende Relevanz von VR-Lernumgebungen [Ba08] erreichen auch virtuelle Klassenzimmer die Lehrerausbildung. Im Folgenden werden einige dieser Systeme vorgestellt.

*TLE TeachLive* [Di07] bietet die Möglichkeit in einem fach-neutralen Unterricht Routinen in einem Klassenzimmer mit 5 SchülerInnen zwischen 10 und 15 Jahren zu erproben. Die virtuellen SchülerInnen haben fest eingestellte Charakteristika und interagieren mit dem Lehrenden in einer für einen ruhigen Unterricht angemessenen Art. Als visuelles Medium wird ein Curved Display genutzt. Die sprachliche Interaktion erfolgt über ein Headset. SchülerInnen werden hier in einem Comic-Stil dargestellt.

Mit *Breaking Bad Behaviors (BBB)* [Lu16] werden das prozedurale Wissen im Klassenmanagement und die videounterstützte Reflektion der Studierenden adressiert. SchülerInnen werden fotorealistisch dargestellt und sind optisch in etwa der vierten oder fünften Klassenstufe einzuordnen. Auch das Raumdesign unterstützt vor allem den Eindruck einer Grundschulklasse. Die 24 SchülerInnen werden durch einen externen Coach mittels einer GUI bedient, während Lehramtsstudierende mit einem Head-Mounted Display (HMD) Unterrichtslektionen durchführen.

Das in diesem Beitrag vorgestellte VR-Klassenzimmer hat viele Überschneidungspunkte mit BBB. Dies umfasst die Intention (Klassenmanagement und Reflektion), die Modellierung der Avatare, das Setting zwischen Coach und Studierenden und die durch Fragebögen in einem universitären Kurs erhobenen Benutzerfreundlichkeit. Allerdings gibt es auch zentrale Unterschiede. Zunächst beruht das Verhalten der virtuellen SchülerInnen des hier vorgestellten VR-Klassenzimmers grundlegend auf Techniken der Klassenführung [Ko06], wohingegen in BBB die Störungen der Klasse in die Kategorien *leicht* (hinlegen, auf das Handy sehen), *mittel* (lachen, schlagen) und *schwer* (auf den Stuhl stellen und Handytöne abspielen) eingeteilt werden. Zudem wurde im VR-Klassenzimmer als Laufzeit- und Entwicklungsumgebung Unity[3] statt Unreal[4] verwendet, welches eine geringe Einstiegshürde (auch für Weiterentwicklungen) mit sich bringt und auch die künftige Integration weiterer Mechanismen (z. B. Eye Tracking) vereinfacht. Weiterhin ist die Schülerschaft optisch in der Sekundarstufe zwei und ethnisch divers. Die GUI für den Coach ist außerdem eine Weboberfläche und damit weniger ressourcenintensiv sowie auch remote nutzbar. Weitere Unterschiede finden sich in Details wie der Anordnung der Tische für einen praxisnahen, typischen Frontalunterricht und den Experimentalaufbau für die Evaluation (z. B. Coach-Feedback nach der Trainingsphase statt während dieser).

[3] https://unity.com
[4] https://www.unrealengine.com

Abschließend war eine Anpassung der BBB-Software aufgrund mangelnder Verfügbarkeit nicht möglich.

# 3 Fallstudie: VR-Klassenzimmer

Das an der Universität Potsdam in der Lehrerbildung eingesetzte VR-Klassenzimmer wird im Folgenden hinsichtlich seiner technischen Eigenschaften und Nutzungsmöglichkeiten beschrieben, bevor ein konkreter Einsatz und dessen Evaluierung beschrieben werden.

## 3.1 Prototyp des VR-Klassenzimmers

Die prototypische Anwendung spannt wie in Abbildung 1 dargestellt einen virtuellen Klassenraum auf in dem man eine Unterrichtseinheit gestalten kann.



Abb. 1: VR-Ansicht aus Studierendenperspektive

Für die Verwendung des entwickelten Prototyps wird typisches VR-Setup aus einem leistungsfähigen Rechner sowie einem HMD mit sechs Freiheitsgraden (z. B. HTC Vive[5]) inkl. dazugehörigen VR-Controllern benötigt. Weiterhin ist ein freier physisch-realer Raum von mind. 3x3m² erforderlich (*Play Area*) und eine kabellose Erweiterung für das HMD empfehlenswert. Softwareseitig werden zudem das im VR-Bereich weit verbreitete SteamVR[6] und ein moderner Browser vorausgesetzt.

Abbildung 2 illustriert die grundsätzlichen Nutzungsmöglichkeiten des VR-Klassenzimmers. Trainierende Studierende nutzen das HMD, die Controller und die Kopfhörer um in die virtuelle Umgebung einzutauchen. Der externe Coach nutzt eine

[5] https://www.vive.com
[6] https://steamcommunity.com/steamvr

Webanwendung für die Manipulation des Verhaltens der virtuellen Schüler. Grundsätzlich möglich ist zudem die Betrachtung des virtuellen Geschehens durch die Augen des trainierenden Studierenden (z. B. auf einer Leinwand für ein größeres Auditorium)

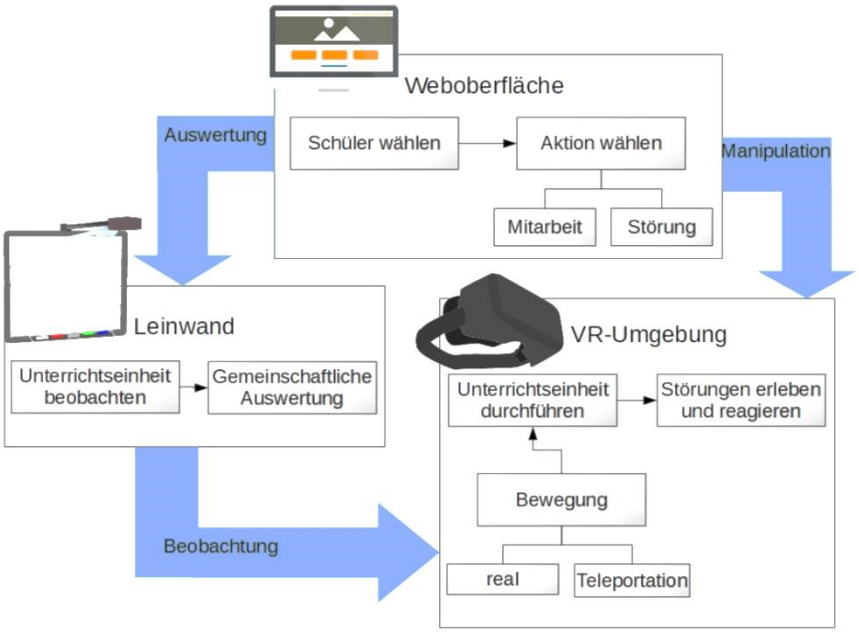



Abb. 2: Nutzungsmöglichkeiten des VR-Klassenzimmers

In der Simulation können Studierende durch reale Positionsveränderung auch ihren virtuellen Avatar durch den virtuellen Raum in gleicher Geschwindigkeit bewegen. Da die reale Begrenzung enger ist als die virtuelle Begrenzung des Klassenraums, ist die Fortbewegung per Teleportation nötig. Hierfür drücken die Studierenden das Touchpad auf den Controllern, zeigen auf einen Punkt zwischen den Bankreihen der SchülerInnen oder vor der Klasse und erscheinen dort nachdem sie das Touchpad wieder losgelassen haben. Für weitere Interaktionsmöglichkeiten, steht es den Studierenden frei ein Buch oder Papierbälle zu greifen mittels des Trigger-Knopfs des Controllers oder den Lichtschalter im Raum zu bedienen durch dessen Berührung mit den Händen des virtuellen Avatars.

Mit Hilfe der in Abbildung 3 dargestellten Webanwendung kann der Coach virtuelle SchülerInnen (einen oder alle) auswählen und bei diesen über ein Dropdown-Menü verschiedene Verhaltensweisen auslösen. Möglich sind dabei unterschiedlich intensive Störungen (z. B. aus dem Fenster schauen, Essen, mit Papierbällen werfen, Nachbarn schlagen) und zwei Formen der Mitarbeit (dem Lehrer mit dem Blick folgen oder Schreiben). Weiterhin kann der Lautstärkepegel über ein weiteres Dropdown-Menü auf drei verschiedene Levels gesetzt werden. Durch diese Bedienung ist es dem Coach möglich, die virtuellen SchülerInnen auf Basis des Verhaltens der StudentInnen zu

steuern. Wenn beispielsweise ein virtueller Schüler mit Papierbällen wirft und der/die Studierende den Schüler ermahnt, nimmt der Coach diese Reaktion wahr, kann den Schüler anklicken und statt des Fehlverhaltens eine Form der Mitarbeit auswählen. Neben dem Frontalunterricht ist derzeit auch Partnerarbeit möglich, bei der die virtuellen Schüler einander zugewandt reden. Andere Arbeitsformen sind derzeit noch nicht implementiert, in der weiteren Entwicklung jedoch angedacht.

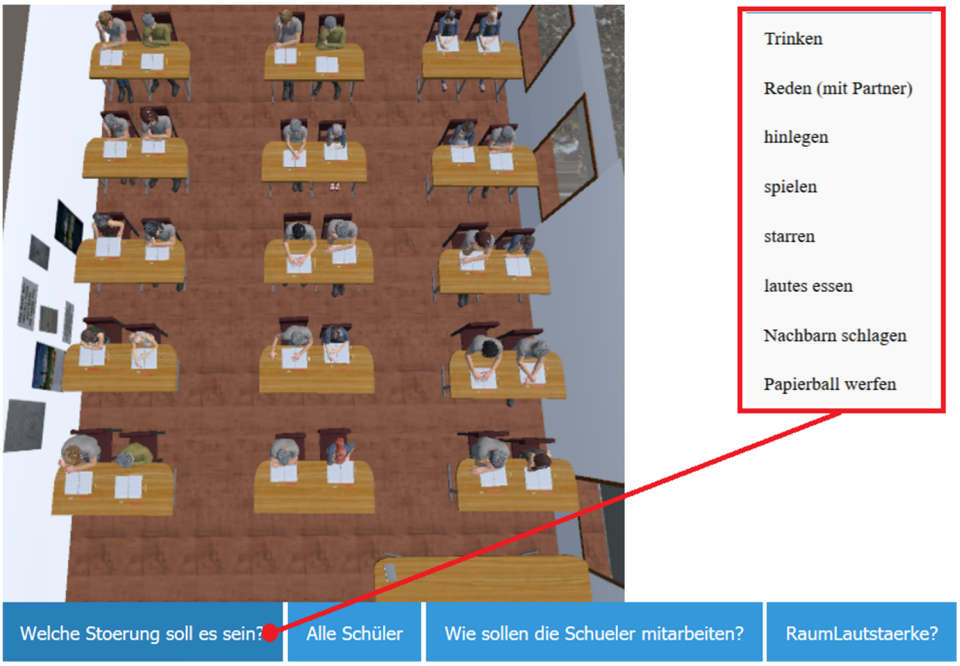



Abb. 3: Webanwendung zur Steuerung des VR-Klassenzimmers durch den Coach

## 3.2 Einsatzszenario: Seminar zum Klassenmanagement

Der Prototyp des VR-Klassenzimmers wurde erstmals im Wintersemester 2018/2019 an der Universität Potsdam im Rahmen eines bildungswissenschaftlichen Seminars zum Thema Klassenmanagement eingesetzt.

Zusammen mit der Instruktion und Motivierung gehört das Klassenmanagement zu den drei übergeordneten Qualitätsdimensionen schulischen Unterrichts und stellt eine zentrale Voraussetzung für den Lernerfolg von SchülerInnen dar [Ku11]. Klassenmanagement meint in diesem Zusammenhang die effiziente Nutzung der zur Verfügung stehenden Unterrichtszeit für Lernprozesse sowie die Vermeidung bzw. Minimierung von Zeitverlusten (z. B. durch Unterrichtsstörungen) [ES01]. Das Klassenmanagement stellt besonders für Novizen eine große Herausforderung dar, da im Studium selten

Gelegenheiten zum Aufbau dieser Kompetenzen bestehen [Lu16]. Aus diesem Grund führt ein Defizit in diesem Bereich nicht selten zu einem erhöhten Belastungsempfinden im Vorbereitungsdienst oder in den ersten Jahren der Berufsausübung. [Fr06].

Vor diesem Hintergrund richtete sich das im Folgenden vorgestellte Seminar an Studierende eines lehramtsbezogenen Bachelors und Masters. Es verfolgte ausgehend von einem kompetenzorientierten Ansatz zwei übergeordnete Ziele: Zum einen sollten Studierende Wissen über theoretische Aspekte und empirische Befunde zur Steuerung von Interaktionsprozessen im sozialen System Schulklasse erwerben. Zum anderen sollte außerdem die Handlungskompetenz der Studierenden erweitert werden, indem sie dieses Wissen in verschiedenen geführten und reflektierten Anwendungssituationen zum Einsatz bringen. Das Seminar untergliederte sich hierfür in vier inhaltliche Teilbereiche:

- Teilbereich 1: Annäherung an die Thematik (Basisdimensionen guten Unterrichts, Konzept des Klassenmanagements, Zusammenhang zwischen Unterrichtsqualität und Klassenmanagement)
- Teilbereich 2: Regeln, Routinen und Verhalten im schulischen Lernen
- Teilbereich 3: Aspekte der Steuerung des Unterrichts (Steuerung des Unterrichtsflusses, Verhaltens- und Aufmerksamkeitssteuerung von SchülerInnen)
- Teilbereich 4: Ursachen von und angemessener Umgang mit Störungen.

Neben der inhaltlichen Erarbeitung der vier Themenbereiche lag der Fokus des Seminars auch auf der Erweiterung der Handlungskompetenz der Studierenden. Zur Förderung dieser situationsspezifischen Fähigkeiten wurden durch den Einsatz von Unterrichtskripten und -entwürfen sowie die Verwendung von authentischen und simulierten Unterrichtsvideos praxisnahe Erprobungssituation konstruiert, die den Studierenden die Möglichkeit boten, das erlernte Wissen anzuwenden und darüber gemeinsam zu reflektieren. Als eine weitere Möglichkeit des authentischen Übens wurde zudem das oben vorgestellte VR-Klassenzimmer eingesetzt.

Das VR-Klassenzimmer wurde im Rahmen des Seminars in zweifacher Weise eingesetzt. Zu Beginn und zum Ende des Seminars gab es zum einen individualisierte Übungsgelegenheiten im virtuellen Klassenzimmer. Dabei befand sich jeweils ein/e Teilnehmer/in des Seminars gemeinsam mit den beiden Dozierenden der Lehrveranstaltung in einem vorbereiteten VR-Labor. Die Studierenden erhielten die Aufgabe, einen vorbereiteten Lehrvortrag zu halten und die aktive Lernzeit durch den Einsatz von Strategien des Klassenmanagements möglichst hoch zu halten. Während die Studierenden ihren Lehrervortrag hielten, lief innerhalb des virtuellen Klassenzimmers ein standardisiertes Skript ab, durch das verschiedene Handlungen der Avatare ausgelöst wurden. Hierbei traten auch Unterrichtsstörungen auf (z. B. “Quatschen”), die in Abhängigkeit des Verhaltens der angehenden Lehrkraft beendet oder fortgesetzt wurden. Nachdem die Studierenden ihren Lehrervortrag im virtuellen Klassenzimmer beendet hatten, wurde die Performanz in einem gemeinsamen Beratungsgespräch mit den Dozierenden evaluiert. Hierbei erhielten die Studierenden zunächst die Möglichkeit, Stärken und Entwicklungspotenziale ihrer Leistung selbst zu identifizieren. Zu dieser

Einschätzung gab es anschließend eine Rückmeldung durch die Dozierenden. Abschließend wurden gemeinsam mögliche Entwicklungsschritte festgehalten.

Neben den individualisierten Übungen wurde zum anderen auch eine Seminargruppensitzung im VR-Labor durchgeführt, um die im Seminar erworbenen Kompetenzen anzuwenden und gemeinsam zu reflektieren. Im Rahmen dieser Lehrveranstaltung erhielten insgesamt zwei Studierende nacheinander die Möglichkeit, ihre im Seminar erworbenen Kenntnisse im VR-Klassenzimmer zu erproben. Ihre Aufgabe bestand darin, eine Partnerarbeitsphase im VR-Klassenzimmer zu überwachen und Strategien des Klassenmanagements anzuwenden, um die aktive Lernzeit hoch zu halten. Der Blick ins VR-Klassenzimmer wurde den anderen SeminarteilnehmerInnen per Leinwand angezeigt und ermöglichte somit allen TeilnehmerInnen an der Erfahrung des Probanden/der Probandin teilzuhaben. Die BeobachterInnen erhielten außerdem verschiedene Beobachtungsaufträge. Hierzu zählen das Identifizieren kritischer Momente in der Lehrer-Schüler-Interaktion sowie die Bewertung des Verhaltens der Lehrkraft in verschiedenen Dimensionen des Klassenmanagements. Im Anschluss an die Übung fand ein gemeinsamer Austausch mit Elementen des Peer-Feedbacks zwischen der Probandin/dem Probanden und den BeobachterInnen statt. Wiederum erhielten zunächst die Probanden die Möglichkeit, ihre Leistung im VR-Klassenzimmer einzuschätzen. Anschließend erhielten die BeobachterInnen die Gelegenheit, Rückmeldung zum gezeigten Verhalten zu geben. Mithilfe einer Screen-Aufnahme war es ferner möglich, einzelne kritische Momente in der Lehrer-Schüler-Interaktion wiederholt zu betrachten und in einen gemeinsamen Austausch darüber zu kommen. Wie die Studierenden ihre Erfahrungen sowohl mit den verschiedenen Lerngelegenheiten des Seminars als auch mit dem virtuellen Klassenzimmer eingeschätzt haben, soll im Folgenden betrachtet werden.

## 3.3 Evaluierung

### Studiendesign und Stichprobe

Im Wintersemester 2018/19 wurden zwei einheitlich konzipierte Seminare an der Universität Potsdam unterrichtet, die im bildungswissenschaftlichen Bachelormodul „Schultheorie und Bildungsforschung“ sowie im bildungswissenschaftlichen Mastermodul „Schule und Unterricht: Analyse, Entwicklung, Evaluation“ mit dem Titel „Klassenmanagement in Theorie und Praxis“ verortet waren. Die Evaluation des Seminars sollte der Ergründung von drei Forschungsfragen dienen:

1. Wie bewerten die Studierenden die Nutzbarkeit der VR-Umgebung?
2. Wie verändern sich die selbstberichteten Kompetenzen im Klassenmanagement im Seminar?
3. Worauf führen die TeilnehmerInnen die selbstberichteten Veränderungen ihrer Kompetenzen zurück?

Die Evaluation erfolgte mithilfe des in Abbildung 4 illustrierten Prä-Post-Test-Designs, wobei sowohl zu Beginn des Semesters als auch am Ende jeweils schriftliche Befragungen

durchgeführt wurden (vgl. Abb. 1). Die Daten wurden deskriptiv analysiert. Insgesamt nahmen 47 Studierende am Seminar und der ersten Befragung teil. Die Studierenden waren zu diesem Zeitpunkt $M$ = 23,5 Jahre ($SD$ = 3,8) alt und befanden sich im Durchschnitt im 6. Hochschulsemester ($M$ = 6,1 Hochschulsemester; $SD$ = 2,5). 43% der Studierenden waren weiblich. Die Mehrheit der Studierenden (83%) besaß zum ersten Messzeitpunkt keinerlei Vorerfahrung mit VR. 91,5% der Studierenden ($n$ = 43) nahmen zu beiden Messzeitpunkten an den schriftlichen Befragungen teil.

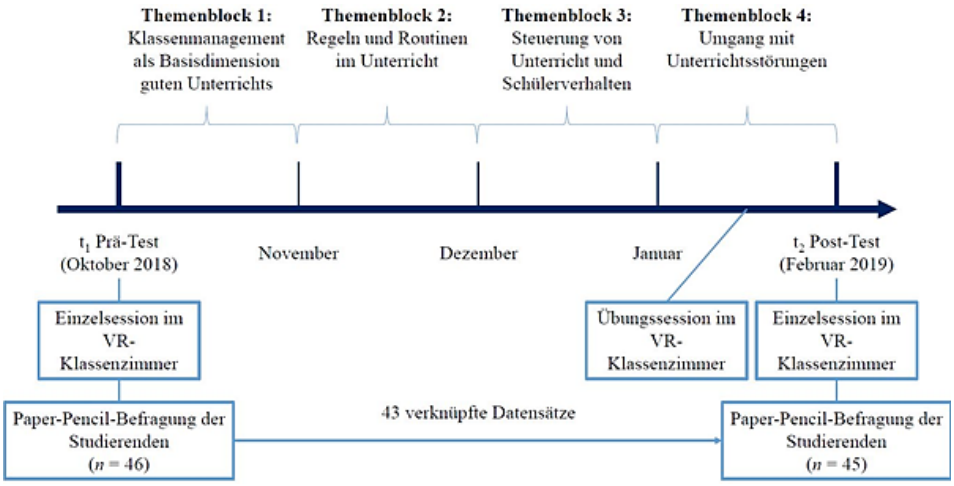



Abb. 4: Studiendesign für die Evaluation des Seminars *Klassenmanagement in Theorie und Praxis*

**Instrumente**

Um die Bewertung der VR-Umgebung zu adressieren, wurde die Bedienbarkeit und Nützlichkeit mithilfe des User Experience Questionnaires (UEQ) [LSH08] bewertet. Im UEQ wird ein persönliches Meinungsbild erhoben zur *Attraktivität*, *Durchschaubarkeit*, *Effizienz*, *Steuerbarkeit*, *Stimulation* und *Originalität* einer Software.

Um die Lernwirksamkeit der Trainingsseminare aus Sicht der TeilnehmerInnen zu erfassen, wurden die Studierenden sowohl zum ersten als auch zum zweiten Messzeitpunkt gebeten, ihre Handlungskompetenz im Klassenmanagement auf einer Skala von 0 (keine Handlungskompetenz) bis 10 (sehr hohe Handlungskompetenz) einzuschätzen. Zusätzlich wurden die Studierenden zum zweiten Messzeitpunkt gebeten, ihre Handlungskompetenz im Klassenmanagement retrospektiv zum ersten Messzeitpunkt einzuschätzen.

Zusätzlich zu der Einschätzung ihrer Handlungskompetenz wurden die Teilnehmenden zum zweiten Messzeitpunkt danach gefragt, worauf sie ihre selbstberichteten Veränderungen zurückführen. Diesbezüglich wurden die Studierenden gebeten für insgesamt zehn verschiedene Lerngelegenheiten (z. B. reale Unterrichtsvideos, theoretischer Input der Dozierenden, Einzelsitzung im virtuellen Klassenzimmer) auf

einer 4er-Likert-Skala anzugeben, wie bedeutsam die jeweilige Lerngelegenheit für den selbstwahrgenommenen Kompetenzzuwachs eingeschätzt wurde (1 = gar nicht bedeutsam bis 4 = sehr bedeutsam).

**Befunde**

Abbildung 5 zeigt die UEQ-Befunde. Die Farbgebung stellt einen Vergleich zu 401 anderen UEQ-Befunden her – überwiegend ohne VR-Bezug. Zu diesem Zweck wurde ein konkretes, bekanntes VR-Lernsystem (*HardDriveExchange* [LZL16]) zum Vergleich hinzugefügt.

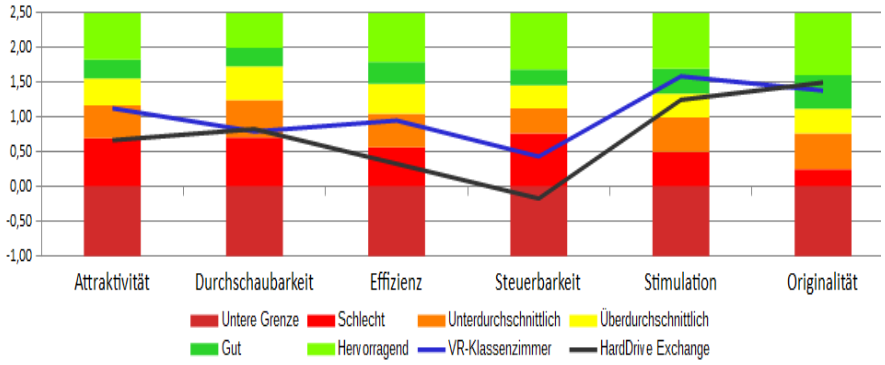



Abb. 5: UEQ-Ergebnisse im Vergleich mit einer anderen VR-Anwendung (HardDriveExchange)

Es zeigt sich die in VR-Anwendungen häufige Stärke in den Dimensionen Stimulation und Originalität. Auch die mangelhafte Steuerbarkeit ist für VR-Anwendungen zunächst nicht ungewöhnlich, da die Nutzer sich an die ungewohnten Bedienkonzepte gewöhnen müssen. Doch auch in den weiteren Dimensionen zeigen sich klar Verbesserungspotentiale für die künftige Weiterentwicklung des VR-Klassenzimmers. So sind bisher beispielsweise nur sehr einfache Interaktionen mit der VR-Welt möglich. Klassenraumtypische Interaktionen wie z. B. das Anfertigen eines Tafelbildes fehlen noch und könnten die Klassenraumsituation authentischer und somit durchschaubarer und effizienter wirken lassen.

Die deskriptiven Ergebnisse zur Frage der selbst berichteten Kompetenzen im Klassenmanagement (Abbildung 6) zeigen einen Anstieg der wahrgenommenen Handlungskompetenz vom ersten zum zweiten Messzeitpunkt. Hierbei handelt es sich um einen substantiellen Lernzuwachs, welcher mit $p < .01$ statistisch signifikant und nach [Co13] praktisch bedeutsam ist. Weiterhin zeigt sich, dass die Studierenden ihr Ausgangsniveau rückblickend geringer einschätzen als zu Beginn des Seminars.

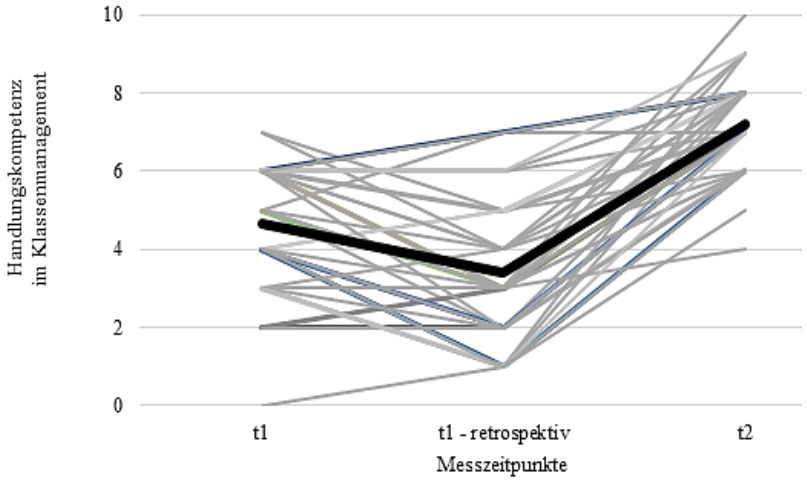



Abb. 6: Selbstberichtete Kompetenzen im Klassenmanagement

Hinsichtlich der Frage, wie bedeutsam die verschiedenen Lerngelegenheiten im Rahmen des Seminars aus Sicht der Studierenden wahrgenommen wurden (Abbildung 7), zeigt sich, dass die Studierenden sowohl die Seminarsitzungen im Allgemeinen, aber auch die Einzel- und Gruppenübungen im virtuellen Klassenzimmer als bedeutsam für ihre Kompetenzentwicklung erlebt haben. Ferner zeigt sich für die verschiedenen Lerngelegenheiten in den Seminarsitzungen, dass sie insbesondere den Einsatz realer Unterrichtsvideos sowie die Arbeit mit Fallbeispielen als bedeutsame Lernerfahrung einschätzen.

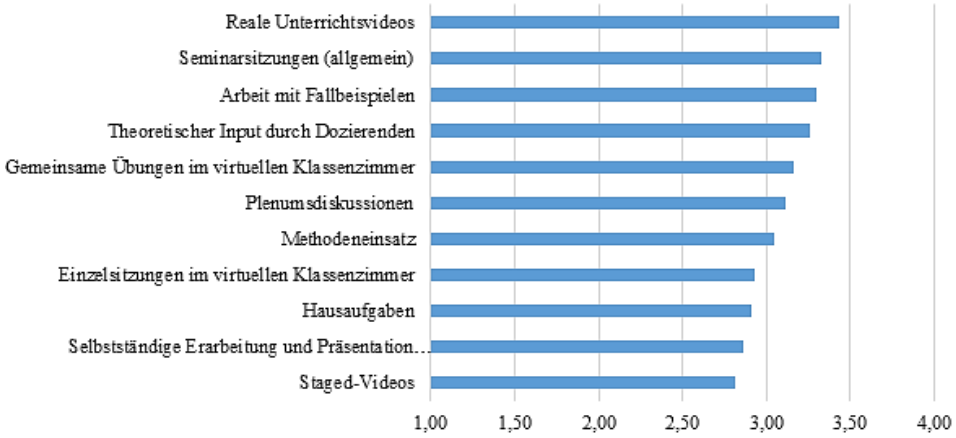



Abb. 7: Bedeutsamkeit der Lerngelegenheiten für den selbstberichteten Kompetenzzuwachs auf einer vierstufigen Likert Skala (1=gar nicht bedeutsam, 4=sehr bedeutsam)

Auch für die kommenden Semester ist eine Durchführung sowie Weiterentwicklung des Seminars geplant. Angedacht ist insbesondere eine intensivere Nutzung des VR-Klassenzimmers. Während dieses bisher für globale Übungen genutzt wurde (z. B. Lehrervortrag halten), könnten zukünftig mehrere kleinere und voneinander isolierte Anwendungssituationen (z. B. Einsatz von Lob) erprobt werden. Die durch den UEQ angedeuteten Schwächen bei der Nutzbarkeit des Systems werden in künftigen technischen Weiterentwicklungen gezielt adressiert und optimiert. Darauf aufbauend wird eine Zunahme der Bedeutsamkeit dieser Lerngelegenheit erwartet.

## 4 Zusammenfassung und Ausblick

VR kann immersive Lernsituationen ermöglichen, die in der physischen Realität nicht ohne weiteres zur Verfügung stehen. Dazu gehört das Training künftiger LehrerInnen zur realitätsnahen Vorbereitung auf mögliche Stresssituationen im Klassenmanagement. In diesem Beitrag wurde ein VR-Klassenzimmer beschrieben das in der Hochschullehre eingesetzt und evaluiert wurde. Die Ergebnisse zeigen einen Mehrwert der neuartigen Lerngelegenheit, wenn auch deren Verbesserungsbedarf zur Steigerung ihrer Bedeutsamkeit im Lehramtsstudium.

Insbesondere wird die Nutzbarkeit des Systems in kommenden Arbeiten verbessert werden – z. B. durch das Hinzufügen neuer Interaktionen in der VR-Simulation. Auch weitere didaktische Settings (neben dem klassischen Frontalunterricht) könnten die Nutzbarkeit des VR-Klassenzimmers erhöhen. Weiterhin wird derzeit an in-situ-Messmethoden zur (teil-)automatisierten Erfassung von Lernleistungen gearbeitet. So erlaubt ein Erfassen der Blickrichtung durch Eye-Tracking beispielsweise den aktuellen Fokus der Studierenden innerhalb der simulierten Klasse zu bestimmen und ggf. zu diskutieren. Darauf aufbauend sind sogar KI-gestützte Reaktionen der simulierten Klasse auf LehrerInnen-Handlungen möglich. Diese Weiterentwicklungen werden begleitend zum fortgesetzten Einsatz des VR-Klassenzimmers in den kommenden Semestern an der Universität Potsdam stattfinden.

Literaturverzeichnis

[Ba08] Bailenson, J. et.al.: The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. Journal of the Learning Sciences, S.102-141, 2008.

[Co13] Cohen, J: Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Taylor and Francis, Hoboken, 2013

[Di07] Dieker, L. et al.: Virtual classrooms: STAR simulator building virtual environments for teacher training in effective classroom management. New Learning Technology SALT, 4, S. 1-22, 2007.

[ES01] Emmer, E.T.; Stough, L. M.: Classroom Management. A Critical Part of Educational Psychology, With Implications for Teacher Education. Educational Psychologist 2/36, S. 103-112, 2001.

[Fr06] Friedman, I.A.: Classroom Management an Teacher Stress and burnout. In (Evertson, C.M. & Weinstein, C.S., Hrsg.): Handbook of classroom management. Research, practice, and contemporary issues. NY: Routledge. New York, S. 925-944, 2006.

[Ko06] Kounin, J.S.: Techniken der Klassenführung. Standardwerke aus Psychologie und Pädagogik. Reprints. Waxmann, Münster, 2006.

[Ku11] Kunter, M. et al.: Die Entwicklung professioneller Kompetenz von Lehrkräften. In (Kunter, M., Baumert, J., Blum, W. & Neubrand, M. Hrsg.): Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV. Waxmann. Münster, S.55-68, 2011.

[LSH08] Laugwitz, B.; Schrepp, M.; Held, T.: Construction and evaluation of a user experience questionnaire, 2008.

[LZL16] Lutze, F.; Zender, R.; Lucke, U.: HardDrive Exchange - Eine VR-Lernanwendung zur Durchführung von Festplattenwechseln in Speichersystemen. In (Lucke, U.; Schwill, A.; Zender, R. Hrsg.): DeLFI 2016 – Die 14. E-Learning Fachtagung Informatik, Bonn. Gesellschaft für Informatik e.V., S. 179-190, 2016.

[Lu16] Lugrin, J.-L. et al.: Breaking Bad Behaviors. A New Tool for Learning Classroom Management Using Virtual Reality. Frontiers in ICT 3/26, 2016.

[PSE15] Paulicke, P.; Schmidt, T.; Ehmke, T.: Hier werden Parallelwelten im Unterricht sichtbar – Multiperspektivische Unterrichtsvideos in der universitären LehrerInnenausbildung. Seminar – A Journal of Germanic Studies 3/15, S. 15-27, 2015

[WP18] Wiepke, A.; Paulicke, P.: Aufbau von Handlungswissen zum Umgang mit Unterrichtsstörungen im VR-Klassenraum. Ausgezeichneter Beitrag im Rahmen des Wettbewerbs „Gelungene VR/AR-Lernszenarien“, VR/AR-Learning, 2018. Online: https://www.uni-potsdam.de/vrarl/downloads/VRARL_Wettbewerb2018-Klassenraum-20180522-final.pdf

[Ze18] Zender, R. et al.: Lehren und Lernen mit VR und AR – Was wird erwartet? Was funktioniert?. In: Proceedings der Pre-Conference-Workshops der 16. E-Learning Fachtagung Informatik (DeLFI 2018), CEUR-WS.org, 2018.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 145

# Wirkung von interaktiven 3D-360°-Lernvideos in der praktischen Ausbildung von Handwerkern

## Vergleich von 3D-360°-Lernvideos mit konventionellen Lernvideos in Bezug auf den praktischen Lernerfolg auf einer Lehrbaustelle

Anna Klingauf[1], Johannes Funk[1], Angela Lüüs[1] und Ludger Schmidt[1]

**Abstract:** Die vorliegende Studie vergleicht den Einsatz von 3D-360°-Lernvideos mit konventionellen Lernvideos bei der praktischen Ausbildung von Handwerkern. Es wurde die Hypothese geprüft, dass Motivation und Ablenkung durch die Technologie einen gegensätzlichen Einfluss auf den Lernerfolg haben und so die vorliegenden gemischten Befunde in diesem Feld erklären. Insgesamt 20 Auszubildende des ersten Lehrjahrs der Ausbildung zum Hoch- bzw. Tiefbaufacharbeiter lernten entweder mit konventionellen Lernvideos am Laptop oder mit 3D-360°-Videos, die auf einem Smartphone-basierten Head-Mounted Display angeschaut wurden. Der Lernerfolg wurde anhand einer praktischen Aufgabe gemessen, die von zwei unabhängigen Ratern bewertet wurde. Die durch die Lernmethode ausgelöste Motivation und Ablenkung der Lernenden wurde mithilfe von Fragebögen gemessen. Wie erwartet zeigte sich in der Gruppe der mit 3D-360°-Videos Lernenden sowohl eine höhere Motivation als auch eine höhere Ablenkung als in der Kontrollgruppe. Der Lernerfolg unterschied sich nicht zwischen den Gruppen. Die erwarteten Korrelationsmuster zur Unterstützung der weiteren Teilhypothesen konnten nicht gefunden werden.

**Keywords:** Virtual Reality, Augmented Reality, 3D-360°-Video, Blicksteuerung, Handwerk

## 1 Einleitung

Die zunehmende Digitalisierung bietet auch mittelständischen Unternehmen innovative Chancen und mögliche Wettbewerbsvorteile [De16]. Allerdings weisen besonders kleine Unternehmen mit weniger als 50 Mitarbeitern häufig Defizite in der Digitalisierung auf; Gründe hierfür sind beispielsweise Unsicherheiten im Datenschutz, hohe Investitionskosten, mangelnde Kompetenzen der Beschäftigten oder fehlende Informationen über Anwendungsmöglichkeiten [SVS16].
Das diesem Beitrag zugrundeliegende Projekt FachWerk verfolgt das Ziel, die Digitalisierung im Handwerk durch die Konzeption und Umsetzung einer bedarfsgerechten Weiterbildung voranzubringen. Im Rahmen des Projektes wird eine digitale Lernplattform eingesetzt, um die Vorteile von Präsenz- und Selbstlernphasen in einem Blended-Learning-Ansatz zu verbinden. Neben konventionellen Lehrmaterialien wie Texten und

[1] Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstraße 7, 34125 Kassel, A.Klingauf@uni-kassel.de

doi:10.18420/delfi2019_298

Videos werden auch neue, technologisch unterstützte Lernformen wie Virtual Reality und Augmented Reality in der Weiterbildung erprobt.

Die Nutzung von neuen Lernformen ist eine mögliche Antwort auf die Wünsche der Lernenden nach spannenden Lernerlebnissen, die neben der Vermittlung neuen Wissens auch Spaß machen [Zi13]. Die in dieser Studie verwendeten 3D-360°-Videos bieten im Vergleich zu herkömmlichen Videos eine zusätzliche Interaktionsmöglichkeit, da die Nutzer*innen nicht nur einen bei der Aufnahme vorbestimmten Bildausschnitt sehen, sondern frei in alle Richtungen schauen können. Für ein immersiveres Erlebnis kann zum Anschauen dieser Videos ein Head-Mounted Display (HMD) genutzt werden, das die Einwirkungen der Außenwelt minimiert [Do17]. Die hohe Immersion einer Anwendung soll dazu führen, dass der virtuelle Ausflug an einen anderen Ort als eigenes Erlebnis wahrgenommen wird und das so Gelernte weniger schnell vergessen wird [Ki17]. Im Gegensatz zu aufwendig animierten VR-Anwendungen werden zur Erstellung von 3D-360°-Videos keine 3D-Animationskenntnisse benötigt. Stattdessen werden reale Situationen mittels spezieller 3D-360°-Kameras aufgezeichnet, wobei durch die doppelte Anzahl an Linsen auch stereoskopische Aufzeichnungen mit Tiefeneffekt (3D) möglich sind. So sind Anwendungen dieser Art für die Bildungsinstitutionen mit einem geringen Mehraufwand, verglichen mit konventionellen Lernmaterialien, umsetzbar [Ka16].

Die vorgestellten 3D-360°-Videos haben sowohl mit Virtual Reality (VR) als auch mit Augmented Reality (AR) einige Gemeinsamkeiten [Do17]. Das Ausblenden der Umgebung und immersive Eintauchen in eine andere Umgebung ist VR-Anwendungen ähnlich, wobei im Unterschied zu VR real existierende, aufgezeichnete Orte genutzt werden. Die Einblendung von interaktiven Schaltflächen und Schrift in die Umgebung ist dagegen an AR angelehnt. Im Anlehnung an die von Domanski et al. verwendeten Begriffe werden die beschriebenen 3D-360°-Videos in der hier vorgestellten Studie Interactive Immersive Media (IIM) genannt [Do17]. Da sowohl VR als auch IIM als immersive Erfahrungen betrachtet werden können, bietet sich der Vergleich der in dieser Studie genutzten IIM-Anwendung mit bestehenden VR-Anwendungen an.

In der Lernforschung wurden 360°-Videos im Lernkontext bereits untersucht. Harrington et al. fanden, dass die Nutzung dieser Lernform im Vergleich zu traditionellen Lernmaterialien zwar zu einer höheren Einbindung der Lernenden, nicht aber zu besseren Lernergebnissen führt [Ha18]. Ähnliche Ergebnisse zeigen sich auch bei VR-Anwendungen mit 3D-animierten Inhalten [PM18]. Parong und Mayer fanden zwar eine höhere Motivation, aber eine niedrigere Lernleistung beim Lernen mit einer VR-Anwendung gegenüber konventionellen Lernmaterialien [PM18]. Dem gegenüber stehen allerdings auch gegenteilige Befunde. So fand Webster beim Vergleich einer hochimmersiven VR-Anwendung gegenüber konventionellen Lernmaterialien eine signifikant höhere Lernleistung [We16]. Darauf aufbauend verglichen Rupp et al. unterschiedlich immersive VR-Anwendungen (z. B. Smartphone, verschiedene HMD) bezüglich der Lernleistung miteinander [Ru18]. Dabei zeigte sich entgegen der Erwartungen kein Vorteil immersiverer Lernformen [Ru18]. Die unterschiedlichen Befunde in diesem Feld wurden von Jensen und Konradsen in ihrem Review zusammengefasst und in verschiedene An-

wendungsfälle aufgeteilt [JK17]. Dabei diskutieren die Autoren auch Ablenkung durch die immersive Erfahrung als möglichen Wirkmechanismus und verweisen dabei auf die Studien von Alves Fernandes et al. [JK17], [Al16]. Im Rahmen dieser Studien wurde die empfundene Ablenkung beim Spielen eines VR-Spiels jedoch nicht explizit erfasst [Al16].

Die hier vorgestellte Studie setzt an diesem Punkt an und untersucht die Hypothese, dass höhere Motivation und höhere Ablenkung durch den Technologieeinsatz zwei Wirkmechanismen mit gegensätzlichem Einfluss sind, die beim Lernen mit immersiven Anwendungen (hier IIM) zum Tragen kommen. Während die erhöhte Motivation durch IIM-Anwendungen einen positiven Einfluss auf den Lernerfolg hat, führt die erhöhte Ablenkung durch das immersive Erleben zu einer Abnahme der Lernleistung. Das Zusammenspiel dieser Mechanismen liefert dabei einen Erklärungsansatz für die unterschiedlichen Befunde in diesem Feld. Um diese Hypothese zu prüfen, werden drei Teilhypothesen formuliert: Bei der Nutzung einer IIM-Anwendung ist sowohl die Motivation als auch die Ablenkung höher als bei der Nutzung von konventionellen Lernmaterialien (H1). Dabei ist Motivation positiv und Ablenkung negativ mit der Lernleistung korreliert (H2), und die Korrelationen sind für die IIM-Gruppe jeweils stärker positiv bzw. negativ als im konventionellen Fall (H3). Darüber hinaus wird erwartet, dass sich kein signifikanter Unterschied zwischen den Gruppen bezüglich der Lernleistung zeigt.

Eine weitere Besonderheit der vorliegenden Studie ist die bisher wenig untersuchte Stichprobe aus auszubildenden Handwerkern. Blanton und Jaccard argumentieren, dass es wichtig für psychologische Forschung ist, in studentischen Stichproben gefundene Effekte an anderen, speziellen Stichproben zu testen, um die Generalisierbarkeit der Befunde sicherzustellen [BJ08]. Die vorliegende Studie verfolgt diesen Ansatz und bietet die Möglichkeit, den Effekt der erhöhten Motivation einer IIM-Anwendung gegenüber konventionellen Lernmaterialien an einer Stichprobe auszubildender Handwerker zu überprüfen. Das Lernszenario wurde dabei an die Stichprobe angepasst, sodass statt Lernaufgaben zu deklarativem Wissen praktische Tätigkeiten erlernt werden sollen. Das für das Handwerk typische Erlernen eines praktischen Vorgehens und die Bewertung einer dazugehörigen praktischen Arbeit wurden bisher nicht im Zusammenhang mit 360°-Videos und VR-Anwendungen untersucht.

# 2 Methode

## 2.1 Stichprobe

Um die Hypothesen zu prüfen, wurde eine experimentelle Studie in der Lehrbaustelle des Berufsförderungswerks des Handwerks in Korbach durchgeführt. Insgesamt nahmen 20 männliche Auszubildende im Durchschnittsalter von 18,3 Jahren ($SD$ = 3,0 Jahre) teil, die sich im ersten Lehrjahr der Ausbildung zum Hoch- bzw. zum Tiefbaufacharbeiter befanden.

## 2.2 Lerneinheiten

Für den Versuch wurde eine IIM-Lerneinheit erstellt. Das 3D-360°-Video für diese Lerneinheit wurde in einer Halle der Lehrbaustelle aufgezeichnet. In der Lerneinheit wird das Anfertigen der ersten Reihe eines Fliesenspiegels gezeigt; dabei wurden die fachlichen Inhalte an schon existierende und am Berufsförderungswerk verwendete Lernvideos angepasst. Diese bereits vorhandenen Lernvideos wurden als konventionelle Lerneinheit für die Kontrollgruppe verwendet. Das Herstellen des Fliesenspiegels wurde in vier Schritte aufgeteilt. Der erste Schritt, das Anrühren des Mörtels, war in der Vergleichseinheit nicht vorhanden und wurde deshalb als konventionelles Video für die Kontrollgruppe nachgedreht. Der zweite Schritt beschreibt das Anbringen der äußeren Fliesen. Darauf folgt als dritter Schritt das Befestigen eines waagerechten Lots und der restlichen Fiesen. Bei diesem Schritt war in der Vergleichseinheit keine Audioerklärung vorhanden, weshalb die Erklärung für beide Lerneinheiten hinzugefügt wurde. Zuletzt wird im vierten Schritt ein Mörtelkeil über der ersten Fliesenreihe angebracht. In der IIM-Anwendung findet jeder Schritt an einem anderen Arbeitsplatz statt. Die vier Arbeitsplätze sind mit je 90° Versatz um die 3D-360°-Kamera angeordnet, so dass in jeder Blickrichtung jeweils nur ein Schritt zu sehen ist. Der Aufbau ist in Abb. 1 zu sehen. Aufgezeichnet wurde das Video mit einem Auszubildenden aus einem fortgeschrittenen Lehrjahr, der geübt im Anfertigen eines Fliesenspiegels ist. Während der Aufnahme führte er die vier Schritte nacheinander an den jeweils dafür vorgesehenen Arbeitsplätzen durch. So kann die Nutzerin / der Nutzer dem Protagonisten des Videos folgen und verliert nicht die Orientierung im 3D-360°-Video.

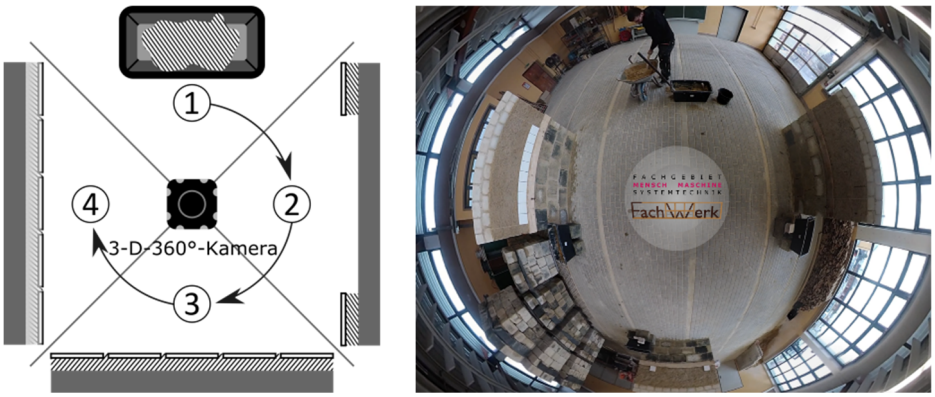



Abb. 1: Aufbau der vier Arbeitsschritte während der Aufzeichnung des 3D-360°-Videos (links); Ausschnitt aus dem aufgezeichneten Video in der Fischaugen-Ansicht (rechts)

In der Nachbearbeitung des 3D-360°-Videos wurde eine Tonspur zur Erklärung der Schritte mit demselben Wortlaut wie in der konventionellen Lerneinheit hinzugefügt. In den Hintergrund des Videos wurden mithilfe des Videobearbeitungsprogramm Adobe Premiere Pro vier Tafeln gelegt, die die Nummer und Beschreibung des gerade sichtbaren Schrittes anzeigen. Zusätzlich wurden mithilfe der Programmierumgebung Unity

Schaltflächen hinzugefügt, mit denen die Anwendung beim Ansehen auf einem HMD durch Kopfbewegungen gesteuert werden kann. Wenn die Nutzerin / der Nutzer die Markierung in der Mitte des Sichtfeldes auf den Button ausrichtet, füllt sich die Markierung farbig aus und löst nach wenigen Sekunden die Funktion der Schaltfläche aus, was auch als Blicksteuerung bezeichnet wird. Die Verwendung dieser Eingabemethode benötigt keine zusätzliche Hardware. Um den Nutzer*innen während des Versuchs die Funktionsweise der Schaltflächen zu vermitteln, wurde eine Einleitungssequenz hinzugefügt. Diese besteht aus einem Standbild der Halle mit den aufgebauten Arbeitsplätzen und einer Audiospur, in der die Funktionsweise des Videos erläutert wird. Um die Einleitungssequenz zu beenden, muss eine Schaltfläche benutzt werden. So kann sichergestellt werden, dass jede Nutzerin / jeder Nutzer die Funktionsweise der Schaltflächen versteht. Die Funktionen der Schaltflächen in der Hauptsequenz umfassen das Starten, Pausieren und das erneute Starten der Hauptsequenz sowie das Überspringen der Einleitungssequenz. Diese Schaltflächen befinden sich auf dem Boden unterhalb der Nutzerin / des Nutzers und damit außerhalb des Sichtfelds. Zusätzlich wurden Schaltflächen, mit denen das Springen zu einem bestimmten Arbeitsschritt möglich ist, hinzugefügt. Die Schaltflächen befinden sich unter den Tafeln mit den Bezeichnungen der Schritte und ermöglichen es, einzelne Schritte beliebig oft in beliebiger Reihenfolge anzusehen. In Abb. 2 ist der erste Arbeitsschritt mit der Tafel, die Aktivierung einer Schaltfläche (links) sowie das Videomenü (rechts) zu sehen.




Abb. 2: Sicht der Nutzer*innen für den ersten Arbeitsschritt mit der Schaltfläche zum Neustarten mittels Blicksteuerung (links); Videomenü unterhalb der Nutzerin / des Nutzers (rechts)

Ziel der Aufnahme und Bearbeitung war, die Inhalte der Vergleichseinheit und der IIM-Anwendung möglichst gleich zu gestalten. Die Arbeitsschritte und Techniken sowie die Audioerklärungen der beiden Anwendungen sind identisch. In der Vergleichseinheit werden einzelne Techniken in Detailaufnahmen gezeigt, was bei 3D-360°-Videos nicht möglich ist, da die Kameraposition festgelegt ist und nicht gezoomt werden kann.

## 2.3 Messung

Zur Messung des Lernerfolgs wurde die Bewertung der praktischen Ausführung der Aufgabe, also das Legen des Fliesenspiegels, herangezogen. Die Bewertung wurde von zwei unabhängigen Ratern erstellt, die keine Information über die Gruppenzugehörigkeit der Probanden hatten. Die systematische Bewertung wurde anhand eines vorher festgelegten Schemas erstellt, das sieben unterschiedlich gewichtete Kategorien enthielt. Die Kategorien „Erkennen der Aufgabenstellung", „Arbeitsablauf" und „Werkzeuge und Material" konnten mit 1-5 Punkten bewertet werden, die Kategorien „Lot und Waage", „Flucht und Winkel", „Maße" und „Gesamteindruck" dagegen mit 1-20 Punkten, woraus sich ein Gesamtwert von 7-95 Punkten ergab. Für den Gesamtwert wurde der Mittelwert aus den Bewertungen der beiden Rater gebildet.

Zur Messung von Motivation und Ablenkung wurden zwei Skalen aus jeweils drei speziell für den Versuchsaufbau formulierten Items gebildet, die auf einer Likert-Skala von 1 (stimme überhaupt nicht zu) bis 5 (stimme voll zu) zu beantworten waren. Die Items sind in Tab. 1 zu sehen. Darüber hinaus wurden die Skalen „subjektiver Lernerfolg" (AL, 3 Items) und „subjektive Meinung" (AO, 4 Items) in den Fragebogen aufgenommen.

| Ablenkung (AD) | Motivation (AM) |
|---|---|
| Ich konnte mich gut auf die Inhalte der Videos konzentrieren. | Wenn hier öfter solche Videos genutzt werden würden, wäre ich motivierter. |
| Ich glaube, dass ich alles Wichtige, was in den Videos gezeigt wurde, gesehen habe. | Die Videos haben dafür gesorgt, dass ich weniger Lust habe, den Fliesenspiegel herzustellen. (inv.) |
| Ich wurde durch die Technologie von den Inhalten der Videos abgelenkt. (inv.) | Die Videos haben meine Motivation für die Aufgabe verstärkt. |

Tab. 1: Skalen für Ablenkung und Motivation mit den dazugehörigen Items: Werte der Items mit dem Vermerk „inv." wurden in der Auswertung invertiert.

## 2.4 Apparatus

Zum Ansehen der IIM-Anwendung wurde eine Zeiss-VR-One-Plus-Halterung genutzt. In diese verstellbare Halterung wird ein Smartphone eingelegt und so als HMD verwendet (siehe Abb. 3, links). Die Lagesensoren des Smartphones reagieren auf die Kopfbewegungen und steuern den Bildausschnitt in der IIM-Anwendung. Bei den Smartphones handelte es sich um ein BQ Aquaris X und ein OnePlus X. Diese beiden Smartphones wurden ausgewählt, da ihre Displays eine identische Auflösung von 1080 x 1920 Pixeln aufweisen und mit 5,2" bzw. 5,0" eine vergleichbare Displaygröße haben. Das Ansehen der konventionellen Lerneinheit geschah auf zwei 14"-Laptops von Lenovo (T440, L420) mit einer Auflösung von je 1368 x 768 Pixeln, auf denen ein Ordner mit den drei

Videodateien geöffnet war. Diese konnten mit dem Windows-Media-Player angesehen werden.



Abb. 3: Vorbereitung im Versuchsraum mit HMDs oder Laptops (links); praktische Umsetzung des Erlernten in einer Halle der Lehrbaustelle unter Beobachtung eines Raters (rechts)

### 2.5 Versuchsablauf

Jeweils vier Probanden wurden zufällig ausgewählt und in den Versuchsraum gebeten. Dort wurden ihnen zufällig eine Teilnehmernummer sowie eine Gruppennummer zugeordnet. Die Gruppennummer entschied darüber, ob sie das konventionelle Lernvideo auf dem Laptop oder die IIM-Lernanwendung zur Vorbereitung auf die praktische Arbeit nutzen sollten. Anschließend sahen die Probanden die Anwendung, wobei alle Probanden Kopfhörer trugen. Für das Nutzen der Anwendungen gab es keine Zeitbegrenzung. Nach dem Ansehen füllten die Probanden die Fragebögen aus. Der letzte Versuchsteil bestand darin, die praktische Aufgabe zu erfüllen. Dafür fertigten die Probanden den Fliesenspiegel in einer Halle der Lehrbaustelle an vorbereiteten Arbeitsplätzen an (siehe Abb. 3 rechts). Dabei wurden sie von zwei Ratern beobachtet. Während der Durchführung wurden den Probanden keine Fragen beantwortet, und es gab keine Zeitbegrenzung.

## 3 Ergebnisse

Zu Beginn der Auswertung wurden alle Skalen auf Normalverteilung geprüft. Lediglich die selbst erstellte Skala „subjektiver Lernerfolg“ weist keine Normalverteilung auf, sondern ist linksschief. Zur Validierung der selbsterstellten Skalen wurde die interne Konsistenz mithilfe von Cronbachs Alpha und das zugehörige 95-%-Konfidenzintervall (KI) berechnet. Dabei ergab sich für die Skala Ablenkung (AD) ein Wert von $\alpha = 0{,}63$ (KI = 0,35-0,91) und für die Skala Motivation (AM) ein Wert von $\alpha = 0{,}62$ (KI = 0,33-0,91). Für die zwei weiteren, hier nicht näher erläuterten Skalen, lagen die Werte im gleichen Bereich.

Ein Welch t-Test zeigt beim Lernerfolg entsprechend der Erwartungen keinen signifikanten Unterschied zwischen der IIM-Gruppe ($M = 79{,}27$, $SD = 10{,}49$) und der Kontrollgruppe mit konventionellen Lernmaterialien ($M = 80{,}50$, $SD = 6{,}13$), $t(14{,}51) = -0{,}33$, $p = 0{,}63$, $1\text{-}\beta = 0{,}09$, siehe Abb. 4 (links). Explorativ wurde auch die Meinung der Lernenden untersucht. Die subjektive Meinung der Lernenden (AO) zu der von ihnen benutzten Lernmethode unterschied sich bei einseitiger Testung signifikant zwischen der IIM-Gruppe ($M = 4{,}28$, $SD = 0{,}56$) und der Kontrollgruppe ($M = 3{,}38$, $SD = 0{,}64$), $t(17{,}69) = 3{,}36$, $p < 0{,}01$.

Teilhypothese H1 besagt, dass sowohl die Motivation als auch die Ablenkung in der Gruppe der mit der IIM-Anwendung Lernenden höher ist als in der Gruppe, die mit konventionellen Materialien lernt. Zur Prüfung wurde jeweils ein einseitiger Welch t-Test für unabhängige Stichproben berechnet. Die mit der Skala AM durchschnittlich gemessene Motivation ist in der IIM-Gruppe ($M = 3{,}80$, $SD = 0{,}69$) höher als in der Kontrollgruppe ($M = 3{,}13$, $SD = 0{,}63$), siehe Abb. 4 (Mitte). Dieser Unterschied ist signifikant $t(17{,}87) = 2{,}25$, $p < 0{,}05$, womit sich eine mittlere Effektstärke von $r = 0{,}47$ ergibt. Die durchschnittliche Ablenkung (AD) ist ebenfalls in der IIM-Gruppe ($M = 2{,}17$, $SD = 0{,}96$) höher als in der Kontrollgruppe ($M = 1{,}57$, $SD = 0{,}35$), siehe Abb. 4 (rechts). Dieser Unterschied ist ebenfalls signifikant $t(11{,}40) = 1{,}86$, $p < 0{,}05$, womit sich eine mittlere Effektstärke von $r = 0{,}48$ ergibt.

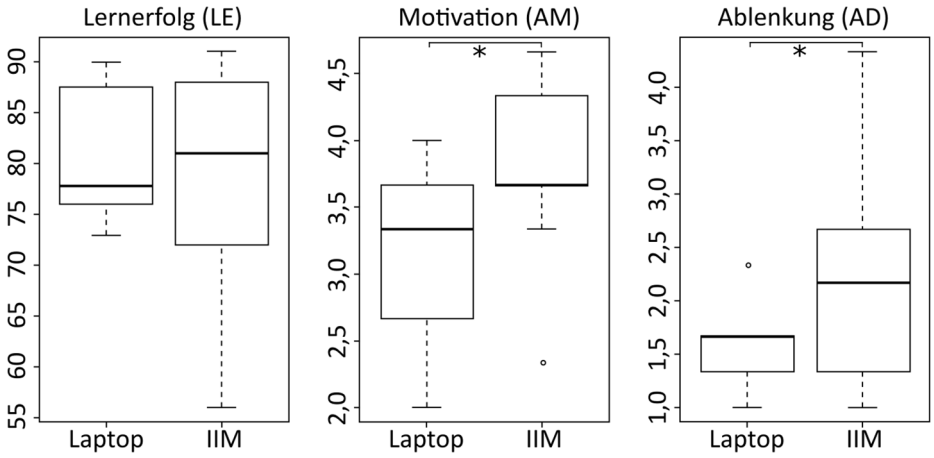



Abb. 4: Boxplots der Skalen Lernerfolg, Motivation und Ablenkung, * p < 0,05

Teilhypothese H2 besagt, dass eine positive Korrelation zwischen Motivation und Lernerfolg und eine negative Korrelation zwischen Ablenkung und Lernerfolg besteht. Dazu wurden jeweils Pearsons Produkt-Moment-Korrelationen zwischen den Variablen berechnet und einseitig auf Signifikanz getestet. Bei nicht signifikanten Korrelationskoeffizienten wird zusätzlich die mit G*Power berechnete Power ($1\text{-}\beta$) berichtet [Fa09]. Über beide Gruppen hinweg beträgt die Korrelation zwischen Motivation und Lernerfolg $r = 0{,}19$ und unterscheidet sich bei einseitiger Testung nicht überzufällig von null

($p$ = 0,21, $1\text{-}\beta$ = 0,21). Die Korrelation zwischen Ablenkung und Lernerfolg beträgt über beide Gruppen hinweg $r$ = -0,35 und unterscheidet sich ebenfalls nicht signifikant von null ($p$ = 0,06, $1\text{-}\beta$ = 0,49). In Abb. 5 sind zusätzlich die Korrelationen der selbst erstellten Skalen untereinander und mit dem Lernerfolg zu sehen.

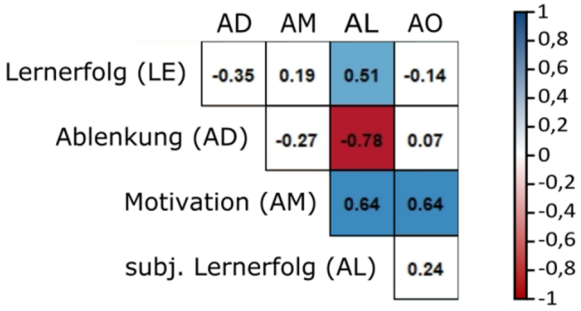



Abb. 5: Korrelationen der Skalen über beide Gruppen hinweg, farblich markierte Felder enthalten einseitig signifikante Korrelationen (p < 0,05), Skala AO: subjektive Meinung

Teilhypothese H3 besagt, dass die Korrelation von Motivation und Lernerfolg für die IIM-Gruppe stärker positiv und die Korrelation von Ablenkung und Lernerfolg stärker negativ ist als für die Kontrollgruppe. Um diese Hypothese zu prüfen, wurde erst für jede Gruppe getrennt Pearsons Produkt-Moment-Korrelation zwischen den Skalen inklusive eines einseitigen Signifikanztests berechnet, und anschließend wurden die beiden Korrelationen mithilfe der z-Statistik verglichen. Der Unterschied zwischen der Korrelation von Motivation mit Lernerfolg in der IIM-Gruppe ($r$ = 0,41, $p$ = 0,51, $1\text{-}\beta$ = 0,22) und in der Kontrollgruppe mit konventionellen Lernmaterialien ($r$ = -0,03, $p$ = 0,53, $1\text{-}\beta$ = 0,22) beträgt $\Delta z = 0.89$ und ist bei einseitiger Testung nicht signifikant unterschiedlich von null ($p$ = 0,19, $1\text{-}\beta$ = 0,23). Der Unterschied zwischen der Korrelation von Ablenkung und Lernerfolg in der IIM-Gruppe ($r$ = -0,53, $p$ = 0,06, $1\text{-}\beta$ = 0,56) und der Korrelation der beiden Skalen in der Kontrollgruppe ($r$ = 0,44, $p$ = 0,90, $1\text{-}\beta$ = 0,41) beträgt $\Delta z = 1{,}98$ und ist signifikant unterschiedlich von null ($p$ < 0,05).

## 4 Diskussion

Die technische Weiterentwicklung im Bereich der Head-Mounted Displays macht die Nutzung von VR-Anwendungen und IIM-Anwendungen im Kontext des Lernens zunehmend interessanter [JK17]. Es ist jedoch wichtig, die Vor- und Nachteile dieser Technologien genauer zu untersuchen, um die unterschiedlichen Befunde in diesem Feld besser zu verstehen [JK17]. Die vorliegende Studie untersuchte deshalb die Hypothese, dass beim Lernen mit immersiven Anwendungen (VR oder IIM) höhere Motivation und höhere Ablenkung den Lernerfolg in entgegengesetzte Richtungen beeinflussen. Die Ergebnisse der Auswertung von Teilhypothese H1 zeigen, dass die IIM-Gruppe gegenüber der Kontrollgruppe sowohl höhere Motivation als auch höhere Ablenkung durch die

Lernmethode aufweist. Diese Befunde entsprechen den Erwartungen und sind auch konsistent mit den bisherigen Forschungsergebnissen, wie beispielsweise von Parong und Mayer, die eine höhere Motivation durch die VR-Anwendungen beschreiben [PM18]. Bei der Prüfung der zweiten Teilhypothese H2 konnten hingegen keine signifikanten Korrelationen gefunden werden. Die berechneten Korrelationskoeffizienten weisen allerdings sowohl für die Korrelation von Motivation mit Lernerfolg, als auch für die Korrelation von Ablenkung mit Lernerfolg eine Tendenz in die erwartete Richtung auf. Die geringe Stichprobengröße von N = 20 führt darüber hinaus zu einer geringen Teststärke, die ebenfalls dafür sorgen könnte, dass der Unterschied der Korrelationen von null nicht signifikant ist. Explorativ lässt sich beobachten, dass die erwarteten Korrelationsmuster Signifikanz erreichen, wenn man statt des bei der praktischen Aufgabe gemessenen Lernerfolgs den subjektiv eingeschätzten Lernerfolg heranzieht. Sowohl die Korrelation zwischen Motivation und subjektivem Lernerfolg ($r = 0{,}64$) als auch die Korrelation zwischen Ablenkung und subjektivem Lernerfolg ($r = -0{,}78$) sind signifikant von null unterschiedlich und entsprechen den Erwartungen. Diese Beobachtung könnte einerseits statistisch darauf zurückzuführen sein, dass Korrelationen zwischen Maßen aus der gleichen Quelle wegen ähnlicherer Fehlervarianzen höher sind als zwischen Maßen aus verschiedenen Quellen. In diesem Fall könnte die Beobachtung als Hinweis gedeutet werden, dass die erwarteten Korrelationen vorliegen und wegen mangelnder Teststärke bei unterschiedlichen Quellen nicht gefunden werden konnten. Andererseits könnte diese Beobachtung darauf zurückgehen, dass die Mechanismen Ablenkung und Motivation vor allem den subjektiven Lernerfolg beeinflussen und dieser wiederum den tatsächlichen Lernerfolg beeinflusst. Um diese Frage aufzuklären, ist weitere Forschung nötig.

Im Rahmen der Auswertung zu Teilhypothese H3 wurden die gleichen Korrelationen nach Experimentalgruppen getrennt berechnet, wobei sich ebenfalls keine der Korrelationen signifikant von null unterschied. Auch hier ist die sehr geringe Teststärke aufgrund der geringen Stichprobengröße von nur noch N = 10 zu beachten. Die Korrelation zwischen Motivation und Lernerfolg in der IIM-Gruppe ist nicht signifikant größer als diejenige in der Kontrollgruppe, auch wenn die Tendenz des Unterschieds in die erwartete Richtung weist. Auch hier ergibt die Power-Analyse eine sehr geringe Teststärke. Die Korrelation zwischen Ablenkung und Lernerfolg ist hingegen für die IIM-Gruppe negativ, für die Kontrollgruppe leicht positiv, woraus sich ein signifikanter Unterschied entsprechend der erwarteten Richtung ergibt. Zusammengenommen lassen die Befunde der Studie keine Aussage bezüglich der Gesamthypothese zu, da bei zwei der drei festgelegten Teilhypothesen die Nullhypothese beibehalten werden muss. Der Befund, dass es trotz der Gruppenunterschiede in Motivation und Ablenkung keine Gruppenunterschiede beim Lernerfolg gibt, ist jedoch ein Hinweis, dass die aufgestellte Hypothese nicht unplausibel ist und weiterer Forschung bedarf.

Die beschriebenen Ergebnisse zeigen deutlich das Spannungsfeld, in dem sich Forschende bei der Auswahl der Stichprobe befinden. Wie von Blanton und Jaccard beschrieben, profitiert Forschung von diversen Stichproben und der Untersuchung spezieller Teilpopulationen [BJ08]. Dieser Ansatz kann jedoch zu Einschränkungen bezüglich der Anzahl der zu Verfügung stehenden Versuchspersonen führen. Für die vorliegende Studie konn-

te zwar eine vollständige Klasse aus dem ersten Ausbildungsjahr rekrutiert werden, die Stichprobengröße mit N = 20 ist jedoch trotzdem klein und führt zu den bereits beschriebenen Problemen bezüglich der Teststärke. Die Ergebnisse der Mittelwertvergleiche leisten jedoch einen wichtigen Beitrag zur Forschung bezüglich der Effekte von immersiven Anwendungen im Lernkontext. Sie zeigen, dass sich der Effekt der durch immersive Anwendungen erhöhten Motivation von Studierenden auf Handwerker übertragen lässt und dass auch die Ablenkung im Vergleich zum Lernen mit konventionellen Lernmaterialien steigt. Ein weiterer interessanter Aspekt ist die Generalisierbarkeit der bisherigen Befunde auf komplexe praktische Lernaufgaben. Während sich viele Studien mit dem Erlernen kognitiver oder psychomotorischer Fertigkeiten wie beispielsweise Jonglieren beschäftigen [JK17], überträgt die vorliegende Studie die Ergebnisse auf eine klassische handwerkliche Aufgabe, das Legen eines Fliesenspiegels. Die Befunde legen dabei nahe, dass sich die Lernleistung zwar nicht unterscheidet, die subjektive Meinung der Lernenden jedoch bei IIM-Anwendungen deutlich positiver ist. Daraus ergibt sich für Ausbildungsbetriebe die Chance, ihre Attraktivität mit dem Einsatz von IIM-Lernmethoden zu steigern. Mit der in dieser Studie vorgestellten Aufzeichnung von 3D-360°-Videos und der Umsetzung mit Smartphones als HMD ist die Verwendung von immersiven Anwendungen für einen großen Kreis von Menschen zugänglich, da weder Programmier- oder 3D-Animationskenntnisse noch größere Investitionen nötig sind.

Die offensichtlichste Limitation dieser Studie ist die geringe Stichprobengröße. Da das Handwerk und andere praktische Ausbildungen jedoch ein großes, noch wenig erforschtes Anwendungsgebiet für immersive Lernmethoden darstellen, ist eine Erweiterung der vorliegenden Studie vielversprechend. Auch die Verbesserung der verwendeten Skalen durch das Hinzufügen weiterer Fragen könnte die Messgenauigkeit ähnlicher Untersuchungen in Zukunft verbessern. Um die Hypothese der gegensätzlichen Einflüsse von Motivation und Ablenkung durch immersive Technologien auf den Lernerfolg weiter zu prüfen, ist ebenfalls weitere Forschung mit größeren Stichproben nötig. Diese Forschungsfrage bekommt besondere Bedeutung bei der Überlegung, die Ablenkung durch immersive Technologien durch die Gestaltung der Lerneinheiten zu reduzieren, sodass der Effekt der erhöhten Motivation allein zutage tritt. Während noch viel Forschungsbedarf im Feld Virtual Reality besteht, sind die Ergebnisse in Lernkontexten weiterhin als vielversprechend einzuschätzen.

Das Vorhaben FachWerk (FKZ 02L15A190) wird im Rahmen des Programms „Zukunft der Arbeit“ vom Bundesministerium für Bildung und Forschung und dem Europäischen Sozialfonds gefördert. Wir danken dem Berufsförderungswerk des Handwerks Korbach und der Kreishandwerkerschaft Waldeck-Frankenberg für die Ermöglichung der Studie. Besonderer Dank gilt Herrn Froese, Herrn Emde und Herr Schüttler für die tatkräftige Unterstützung der Planung, Vorbereitung und Durchführung der Studie.

## Literaturverzeichnis


[Al16] Alves Fernandes, L. M. et al.: Exploring educational immersive videogames: an empirical study with a 3D multimodal interaction prototype. Behaviour & Information Technology, 35/16, S. 907–918, 2016.

[BJ08] Blanton, H.; Jaccard, J.: Representing Versus Generalizing: Two Approaches to External Validity and Their Implications for the Study of Prejudice. Psychological Inquiry, 19/08, S. 99–105, 2008.

[De16] Demary, V. et al.: Digitalisierung und Mittelstand. Eine Metastudie. Institut der deutschen Wirtschaft Medien, Köln, 2016.

[Do17] Domanski, M. et al.: Immersive visual media — MPEG-I: 360 video, virtual navigation and beyond. In (IWSSIP Hrsg.): IWSSIP. Proceedings. IEEE, [Piscataway, NJ], S. 1–9, 2017.

[Fa09] Faul, F. et al.: Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavior Research Methods, 41/09, S. 1149–1160, 2009.

[Ha18] Harrington, C. M. et al.: 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention. Journal of Surgical Education, 75/18, S. 993–1000, 2018.

[JK17] Jensen, L.; Konradsen, F.: A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies, 23/17, S. 1515–1529, 2017.

[Ka16] Kavanagh, S. et al.: Creating 360° educational video. In (Duh, H. et al. Hrsg.): Proceedings of the 28th Australian Conference on Computer-Human Interaction - OzCHI '16. ACM Press, New York, New York, USA, S. 34–39, 2016.

[Ki17] King-Thomson, J.: The Benefits of 360˚ videos & Virtual Reality in Education, https://blend.media/blog/benefits-of-360-videos-virtual-reality-in-education, Stand: 23.10.2018.

[PM18] Parong, J.; Mayer, R. E.: Learning science in immersive virtual reality. Journal of Educational Psychology, 110/18, S. 785–797, 2018.

[Ru18] Rupp, M. A. et al.: Investigating learning outcomes and subjective experiences in 360-degree videos. Computers & Education, 128/18, S. 256–268, 2018.

[SVS16] Saam, M.; Viete, S.; Schiel, S.: Digitalisierung im Mittelstand: Status Quo, aktuelle Entwicklungen und Herausforderungen. ZEW-Gutachten und Forschungsberichte. Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, 2016.

[We16] Webster, R.: Declarative knowledge acquisition in immersive virtual learning environments. Interactive Learning Environments, 24/16, S. 1319–1333, 2016.

[Zi13] Ziemann, W.: 360° Projektionen im Bildungsbereich. In (Overschmidt, G.; Schröder, U. B. Hrsg.): Fullspace-Projektion. Mit dem 360°lab zum Holodeck. Springer Berlin Heidelberg, S. 197–211, 2013.

# Virtual Reality as an e-learning tool in the student-centered vehicle development project "proTRon"

Lucas Küntzer[1], Michael Hoffmann[2] and Georg Rock[3]

**Abstract:** Immersive Virtual Reality (VR) is increasingly being used in many engineering disciplines. This paper presents the practical use of an e-learning tool for project-based learning in higher education in the context of a student-centered, industrial-sized project. The benefits and challenges of VR technology in education are examined. The paper briefly outlines the advantages of VR in its practical application in education as well as in the professional industry. In conclusion the paper argues that VR applications can be useful in improving effectiveness and efficiency in existent processes. However, such applications must be designed with great care in order to achieve the desired effects in the long term.

**Keywords:** immersive virtual reality, e-learning, project-based learning, vehicle development

## 1 Introduction

Virtual Reality technology has seen use in simulation, training and visualization for decades. A great number of different applications have been developed for engineering, medicine and computer science [VRL17]. In recent years a lot of progress has been made in VR hardware and software, especially in the field of head-mounted displays. As a result, the over-all relevance of immersive Virtual Reality technology in education continues to rise strongly [VRL17].

This paper outlines the experiences gained with VR technology as an e-learning tool, following a project-based learning approach. It describes the development and use of a VR prototype in the student-centered vehicle development project "proTRon" (cf. [Pr19]). It outlines the potential benefits of similar applications in education as well as in the professional industry. In "proTRon" students are developing the hyper-efficient, future-car concept "proTRon EVOLUTION" (cf. [Pr19]). The goal of our research is to determine to what extent problem-based learning can be complemented by VR within the context of projects like "proTRon". More specifically, we try to find out how to best support the development of problem-solving skills through e-learning in an immersive VR environment.

[1] Trier University of Applied Sciences, Dep. of Computer Science, Schneidershof, 54294 Trier, L.Kuentzer@hochschule-trier.de
[2] Trier University of Applied Sciences, Dep. of Engineering, Schneidershof, 54294 Trier, M.Hoffmann@hochschule-trier.de
[3] Trier University of Applied Sciences, Dep. of Computer Science, Schneidershof, 54294 Trier, G.Rock@hochschule-trier.de

doi:10.18420/delfi2019_265

### 1.1 Related Work

Through the recent technological advancement head-mounted, immersive VR has become a widespread and distinctive way of interacting with a virtual environment, setting itself apart from traditional desktop applications [VRL17].

Competency-based learning has been called an appropriate response to the changing demands of the professional industry [RWB14]. Virtual learning environments can be a suitable framework to teach students competencies, like problem-solving or independence [RWB14]. Interactive, immersive VR can outperform more passive learning methods such as video study or book learning due to its innate active learning approach [AV18]. The advantage of immersive VR lies in its positive impact on emotion and engagement, which is linked to student performance and knowledge confidence [AV18].

To avoid relying on the technology's novelty [AV18], educational VR applications need to be designed with great care, suitable to the target audience and based on the overarching instructional design [Ze18]. While VR in education seems to perform poorly in large groups, it performs especially well in learning environments featuring abstract problems [Ze18], which could be tailored to students' learning preferences for the most optimal outcome [AV18]. As a result, VR tools seem to be a suitable addition for project scenarios and e-learning in higher education [Ma17], a notion which we support in this paper.

## 2 Method

Our institution, as a University of Applied Sciences, puts a focus on teaching students the necessary practical skills and competencies sought for by the professional industry. This means to tackle real-life problems, often in collaboration with companies, and let students work on solutions. Those student projects retain the option to fail as opposed to the need of turning a profit in the professional industry, which allows students to take high risks and pursue creative approaches. This project-based learning approach is deeply ingrained in our Computer Science and Engineering departments and their curriculums. Endeavours like the CAD/CAM-laboratory LDPF (cf. [LD19]) and the interdisciplinary project "proTRon" (cf. [Pr19]) are prominent examples of that.

"proTRon" is a student-driven vehicle development project and most of its sub-projects are graded by professors for course credit. The student team was originally founded to develop vehicles for eco-challenge competitions (cf. [Sh19]) and is now working on the street vehicle concept "proTRon EVOLUTION". Every semester this results in dozens of group projects and theses in a hardware-agile and problem-oriented environment.

In 2016 the LDPF developed a VR presentation for the virtual rollout of the project's next vehicle design. This early tool laid the technical foundation for the "proTRon EVOLUTION VR Experience" (VRXP). The goal of the updated VR application is to

educate and inform users by letting them explore the vehicle’s structure, its key sustainability aspects and the history of the ongoing student project. [Kü18]

The prototypical VRXP was developed as the practical part of a bachelor thesis. An agile development process, featuring frequent development prototypes, short-feedback cycles and a focus on the design phase ensured a high software quality [IS19]. Achieving a high level of usability and maintainability were the main goals coming out of this design phase. This resulted in 54 non-functional requirements and other design artifacts, providing measurable criteria for the prototype to fulfill.

The focus on maintainability meant setting up an effective and efficient pipeline of migrating the 3D data of the vehicle from proprietary CAD software to the employed game engine. It was paramount to be able to retain the original product structure and associated materials, while unifying coordinate systems, fusing little parts like screws and simplifying geometry for performance reasons. The proposed manual pipeline, together with the overall modular design of the software and the proprietary open-source libraries provide an application development framework for the purpose of creating VR applications for use in a product development process. This framework enables the original developer or other students to easily update vehicle parts, information and interactions, provided they act according to the existing documentation. The software’s design facilitates the addition of new modules like new scenarios or additional tutorials, a cornerstone of the prototype’s usability. The application prototype was designed in a way to ease users unfamiliar with VR into the virtual world and teach them how to use the provided tools to explore the VRXP on their own.

## 3 Results

The “proTRon EVOLUTION VR Experience” (VRXP) [Kü18] and its derivatives are highly interactive VR applications with a focus on user exploration. The primary goal of the VRXP is to provide a sandbox environment in which the user can learn more about the “proTRon” project. In its current stage it serves this purpose well, but there is still room for improvement when it comes to ease of access and additional features.

Key features of the “proTRon EVOLUTION VR Experience” include [Kü18]:

- 3D-visualization of detailed, industrial-grade CAD data in the original scale
  - dynamic cross-sections of parts (cf. Fig.1 [2])
  - simulating common real-life interaction patterns (e.g.: doors, steering wheel)
- package space review
  - study of data sheets and background information (cf. Fig.1 [1])
  - package group filtering (cf. Fig.1 [4])

  - manual dis-/assembly of parts (cf. Fig.1 [3,5])
- focus on usability in conjunction with exploratory user-interaction
  - basic “immersive VR” tutorial (e.g.: locomotion, virtual body presence)
  - context-sensitive tutorial prompts for interactions (cf. Fig.1 [5,6])

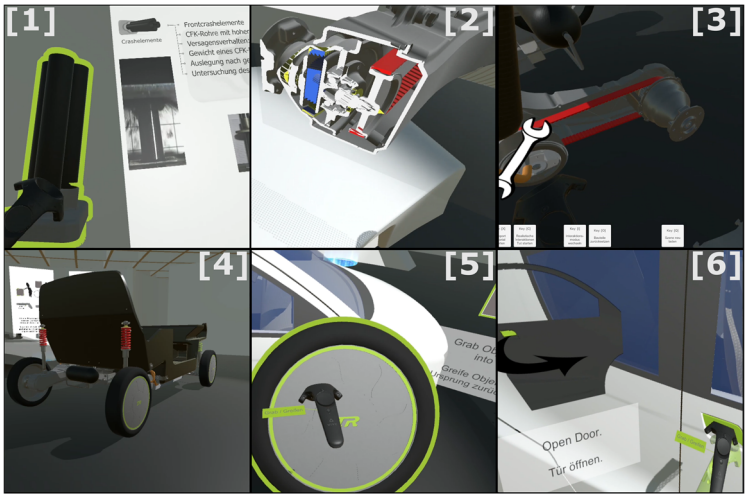



Fig. 1: Visual overview of some interactions and features in the VRXP

Originally the VRXP was showcased at trade expos and visitor events, where users were asked to fill in an evaluation questionnaire regarding the VR experience. The survey focused heavily on the users’ engagement with the virtual world as well as on the perceived usefulness of usability features like the tutorial. Most subjects stated that the VR application caused higher interest in the project “proTRon” and the topic of energy-efficiency. They also reported a higher engagement, compared to traditional media. However, much of that could be attributed to the novelty surrounding the technology. The subjective evaluation by event visitors as well as members of “proTRon” noted the realistic visualization of 3D vehicle data in the original 1:1-scale and the interactive dis-/assembly of parts as the most satisfying and useful features of the application. [Kü18]

Within “proTRon” the VRXP was repeatedly used productively to review design concepts as well as package spaces and to adjust them collaboratively in a live environment. The examination in VR heavily supported the decision-making process through a higher degree of validation, a quicker consensus and a stronger confidence in the decision without the need to spend the time and the resources to build physical prototypes.

In general, VR applications possess a set of characteristics that have the potential to greatly enhance a digitalized product development process (PDP). Aside from the abstract view on product layers and the agility in prototyping that VR applications offer, they can also

enhance or even evolve many traditional processes. Its nature makes VR highly adaptive in its application (cf. Fig.2). In industrial applications VR seems to be particularly strong in the PDP when communicating (interdisciplinary) interfaces, visualizing products, working collaboratively and validating results.

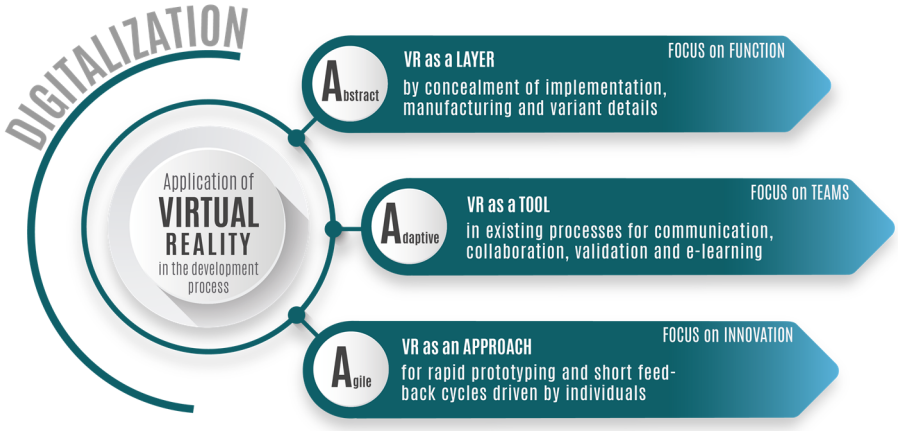



Fig. 2: Three key aspects of Virtual Reality in the development process

Many processes in “proTRon” can be complemented by VR tools, which offer low-cost, reusability and variance exploration. VR as a form of e-learning can succeed in engaging students that have a hard time with traditional learning approaches. VR makes physical products and related problems more accessible to the user, supporting a problem-based learning environment as an e-learning tool. Examples of helpful VR features include visualizing package spaces and designs as well as visually explaining interdisciplinary interfaces or overlaying additional product part information. The technology offers new ways of abstracting complex systems, solving communication difficulties and developing solutions by aid of its immediate, 1:1 scaled visualization. For the students it is rewarding to experience and present their project work in a virtual environment, especially when their designed products are too complicated for immediate manufacturing or prototyping. Within our teaching approach VR has become an additional medium of ever-increasing value, offering a new degree of freedom and creativity in shaping content and facilitating skill gain. In the context of product design, VR is a valuable alternative or supplement to prototyping with physical models or traditional blueprints.

## 4 Conclusion

VR in general seems to have the potential to completely rewrite whole processes in the digitalized industry, being a new way of visualizing and interacting with data. As such we believe that higher education must adjust accordingly and introduce VR technologies to the students. On the one hand it is important to teach the handling of VR applications and

their development, as this is becoming increasingly more relevant in the professional industry. On the other hand, to enable VR as an additional medium in teaching methods is inevitable. In our case that means making VR part of student projects where physical products are being designed and manufactured. Students get to learn how to use VR as a powerful tool for rapid prototyping and validation, helping in the decision-making process and delivering an additional value to the developers by a real-scale visualization.

Going forward, every student team in "proTRon" will have the opportunity to quickly access their data in VR for review and collaboration. The pervasive integration into the design processes of the (sub-)projects is more important than adding new features and improving maintainability and usability of the VRXP.

Our next focus is to integrate the immersive VR technology as a review and prototyping tool into our existing processes and study its effectiveness, acceptance and impact on the student projects. We hope to learn more about how to tailor the medium to support our project-oriented teaching methods in our student-centered industrial-sized environment.

Bibliography

[AV18] Allcoat, D.; Von Muhlenen, A.: Learning in virtual reality: Effects on performance, emotion and engagement. Research in Learning Technology 26: 2140, 2018.

[IS19] ISO/IEC 25010:2011, www.iso.org/standard/35733, accessed: 04/02/2019.

[Je03] Jenny, B.: Projektmanagement, 2003.

[Kü18] Küntzer, L.: Entwicklung einer VR-Anwendung zur Öffentlichkeitsarbeit des Team proTRon. Bachelorabschlussarbeit, Hochschule Trier, 2018.

[LD19] LDPF, hochschule-trier.de/go/ldpf/, accessed: 04/02/2019.

[Ma17] Madathil, K. C. et.al.: An Empirical Study Investigating the Effectiveness of Integration Virtual Reality-based Case Studies into an Online Asynchronous Learning Environment. Computers in Education Journal 8(3), 1-10, 2017.

[Pr19] proTRon, www.hochschule-trier.de/index.php?id=protron, accessed: 04/02/2019.

[RWB14] Reiners, T.; Wood, L. C.; Bastiaens, T.: New Landscapes and New Eyes: The Role of Virtual World Design for Supply Chain Education. Ubiquitous Learning: An International Journal 6, 37–49, 2014.

[Sh19] Shell Eco-Marathon, shell.com/make-the-future/shell-ecomarathon, accessed: 04/02/2019.

[VRL17] Vergara, D.; Rubio, M. P.; Lorenzo, M.: On the Design of Virtual Reality Learning Environments in Engineering. Multimodal Technologies and Interaction 1(2), 11, 2017.

[Ze18] Zender, R. et.al.: Lehren und Lernen mit VR und AR – Was wird erwartet? Was funktioniert?. In (Schiffner, D. ed.): Proceedings of DeLFI Workshops 2018. German Computer Society, Frankfurt, 2018.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 163

# Förderung der Lernmotivation durch adaptives E-Learning

**Komparative Evaluation von Techniken zur adaptiven Nutzerführung**

Mathias Bauer, Jacqueline Schuldt, Beatrix Bau, Marie Jeanne Webers und Heidi Krömker[1]

**Abstract:** Wenngleich Lernmotivation eine wichtige Einflussgröße eines erfolgreichen Lernprozesses darstellt, ist diese bisher kaum Gegenstand der Forschung im Bereich adaptiver E-Learning Systeme. In einem mehrphasigen Studienkonzept soll die Frage beantwortet werden, ob Lernmotivation mithilfe von Adaptation gefördert werden kann. Das aktuelle Papier fokussiert sich daher auf eine komparative Evaluation von drei Adaptationstechniken und deren Einfluss auf die Lernmotivation mit einem Sample von 132 Studierenden in Form einer experimentellen Laboruntersuchung im Vergleich mit einer nicht-adaptiven Version als Kontrollgruppe. Schwerpunkt ist die Auswertung von Logfiles zur Erfassung der Wirksamkeit der Adaptationstechniken und des aktuellen Verlaufs der Lernmotivation während der Arbeit mit einer E-Learning Plattform. Die Adaptation erfolgte auf Basis von Motivations-Selbsteinschätzungen. Die Analyse der Daten gibt erste Rückschlüsse auf die motivationsförderliche Gestaltung der Nutzerführung in E-Learning Systemen.

**Keywords:** Adaptive Nutzerführung, E-Learning, Logfile-Analyse, Lernmotivation

## 1 Einleitung

Motivation wird als wichtige Voraussetzung für eine effiziente Verarbeitung von Lerninhalten und Grundlage für anhaltende Lernfreude angesehen. Das Ziel der vorliegenden Studie ist die Untersuchung der Wirksamkeit von systemseitigen Anpassungen der Nutzerführung als Reaktion auf motivationale Selbsteinschätzungen der Lernenden in einem E-Learning-Setting. Als übergeordnete Forschungsfrage ergibt sich somit: Welche Adaptationstechniken erweisen sich in einer E-Learning Plattform als besonders motivationsförderlich? Das Vorhaben wird in Zusammenhang mit dem vom BMBF geförderten Forschungsprojekt SensoMot realisiert [Sc17]. Grundlage bildet die Lernplattform NanoTecLearn für die Mikro- und Nanotechnik. Diese wurde inhaltlich und softwareergonomisch bereits mehrfach mit Studierenden und Auszubildenden evaluiert und stellt ein solides Fundament für die Erweiterung um adaptive Komponenten dar [Kr16]. Vor der eigentlichen Implementierung wurden zudem bereits mehrere

[1] Technische Universität Ilmenau, Fakultät für Elektrotechnik und Informationstechnik, Fachgebiet Medienproduktion, Gustav-Kirchhoff-Str. 1, 98693 Ilmenau, {mathias.bauer, jacqueline.schuldt, beatrix.bau, marie.webers, heidi.kroemker}@tu-ilmenau.de

doi:10.18420/delfi2019_255

Fokusgruppen mit Studierenden und Experten aus dem Bereich E-Learning-Didaktik, Content, Usability und technischer Umsetzung durchgeführt, um Anforderungen bezüglich Funktionalität und Gestaltung der Adaptationstechniken zu erheben [Ba18]. Nach der Beschreibung theoretischer Grundlagen zu adaptivem E-Learning liegt der Untersuchungsfokus auf der Analyse des Verlaufs der Lernmotivation bei der Arbeit mit drei adaptiven Systemvarianten der Lernplattform. Zu diesem Zweck wurden Logfiles zur Verfolgung der Motivations-Selbsteinschätzungen und adaptiven Navigationspfade durch das Lernmaterial erhoben und ausgewertet. In Verbindung mit weiteren fragebogenbasierten Erhebungen der User Experience und Technologieakzeptanz [Ba19] werden Empfehlungen zur Optimierung der Systemadaptation zur Förderung der Lernmotivation gegeben.

# 2 Theoretische Grundlagen

## 2.1 Adaptives E-Learning

Eine geläufige Kategorisierung im Bereich adaptiver Lernsysteme unterscheidet Makro- und Mikrolevel Adaptation. Diese Differenzierung bezieht sich auf die Häufigkeit der vorgenommenen Systemanpassungen und deren Initiator [Le11]. Angepasst werden können dabei z.B. das User Interface, der Inhalt oder die Benutzerführung eines E-Learning Systems. Die Makrolevel Adaptation ermöglicht Systemanpassungen mit geringer Frequenz, meist einmalig zu Beginn der Nutzung. In diesem Zusammenhang wird auch der Begriff adaptierbares System verwendet [Le11].
Mikrolevel Adaptation geht über diesen Ansatz hinaus, indem das System kontinuierlich Abfragen auf Nutzerseite vornimmt und Anpassungen somit wiederholt durchführbar sind. Grundlage dafür können eine Vielzahl von Nutzer-Charakteristiken bilden, wie z.B. Vorwissen, Ziele oder aktuelle Motivation [So15]. Dieser Ansatz bietet sich insbesondere für die Optimierung der Nutzerführung oder für eine Inhaltsanpassung auf Basis einer Erhebung des Lernfortschritts an. Da aktuell wirksame Motivation idealerweise kontinuierlich erfasst werden sollte, bietet sich diese Form für die vorliegende Studie an. Paramythis et al. [Pa10] beschreiben den Adaptationsprozess für Mikrolevel-adaptive Systeme in fünf Schritten: Die Basis bildet die Erhebung von Nutzerdaten, z.B. die aktuelle Motivation. Daraufhin erfolgt die systemseitige Datenverarbeitung, gefolgt von der Erstellung eines Benutzermodells und Lernhistorie. In Abhängigkeit dieser Datenlage trifft das System eine Entscheidung bezüglich passfähiger Adaptationen. Im letzten Schritt wird eine konkrete implementierte Adaptationstechnik angewendet, um die Anpassung vorzunehmen. An diesem Punkt der Mensch-System-Interaktion wird dem Lernenden die Adaptation bewusst und es kommt zu einer entsprechenden Reaktion auf Nutzerseite. Dies führt wiederum zu Veränderungen der Nutzerzustände und somit zum erneuten Anstoß des Prozesses. Nach der Festlegung auf den übergeordneten Ansatz der Mikrolevel-Adaptation gilt es geeignete konkrete Techniken zur deren systemseitiger Umsetzung zu identifizieren.

## 2.2 Adaptationstechniken

Seit den 1990er Jahren hat sich eine Vielzahl von Adaptationstechniken entwickelt [Kn09], [Sh12], [Gr14]. Brusilovsky unterteilt diese grundlegend in „Adaptive Präsentationstechniken“ und „Adaptive Navigationsunterstützungstechniken“ [Br07]. Adaptive Präsentationstechniken passen den Inhalt an sich an, wohingegen adaptive Navigationsunterstützung dazu dient, den Pfad durch ein E-Learning Systems zu optimieren. Einen umfassenden Überblick über eine Vielzahl konkreter Techniken dieser zwei Kategorien geben Knutov et al. [Kn09]. Eine Adaptation des Inhalts bietet sich vor allem an, wenn dieser in unterschiedlichen Schwierigkeitsgraden vorliegt und eine Systemanpassung auf Basis des Vorwissens vorgesehen ist. Erforderlich dafür sind alternative Inhaltssegmente, Exkurse oder studiengangspezifische Beispiele. Bei adaptiver Navigationsunterstützung geht es hingegen primär um die Anpassung der Sequenzierung der Inhalte auf Basis der Veränderung von Personenfaktoren, wie Vorwissen oder Motivation. Für das aktuelle Vorhaben, welches Anpassungen entsprechend der Lernmotivation vornehmen soll, wurden Techniken der adaptiven Navigationsunterstützung gewählt. Darüber hinaus sollte die bestehende und bereits auf ihre Lernwirksamkeit hin evaluierte Lernplattform NanoTecLearn nicht durch Zusatzinhalte verändert werden. Eine konstant gehaltene inhaltliche Basis ermöglicht zudem das gezielte Manipulieren der Benutzerführung.
Die für die Studie gewählten Techniken der adaptiven Navigationsunterstützung sollten sich insbesondere bzgl. der Freiheitsgrade, die sie für die Benutzerführung bieten, und dem Kontrollempfinden des Lernenden unterscheiden. Diese Kriterien wurden mithilfe vorab durchgeführter Fokusgruppen mit Studierenden identifiziert [Ba18]. Die resultierenden drei Techniken und deren Eigenschaften werden in Tabelle 1 genannt und im nachfolgenden Abschnitt zu Methodik und Studiendesign genauer vorgestellt. Variiert wurde demnach lediglich die Benutzerführung auf Basis der adaptiven Navigationsunterstützungstechniken im Vergleich zur nicht-adaptiven Plattformversion.

| **Technik** | **Link Annotation** | **Direct Guidance** | **Agent** |
|---|---|---|---|
| **Funktionalität** | Hervorheben passender Nachfolgeabschnitte | Geführte Sequenz durch das Lernmaterial | Vorschlag motivationsförderlicher Zusatzinhalte |
| **Kontrollempfinden** | Hoch – Hervorgehobene Links dienen als Empfehlung, Lernender entscheidet selbst, ob diesen gefolgt wird. | Gering – Lernender kann keine anderen Links anwählen und hat keine Wahlfreiheit über Folgeabschnitte. | Hoch – Agent schlägt relevante Zusatzinhalte vor, Lernender kann daraufhin selbst aktiv entscheiden, welchen gefolgt wird. |
| **Freiheitsgrade** | Mittel | Gering | Hoch |

Tabelle 1: Charakteristiken der Treatment-Versionen

# 3 Methodik und Studiendesign

## 3.1 Implementierte Adaptationstechniken

### Nicht-adaptive Version

Die erste implementierte und getestete Version war die nicht-adaptive NanoTecLearn Plattform. Der primäre Wissenszugang erfolgt über text- und bildbasierte Lerneinheiten, die Booklets genannt werden. Dieser Zugang bildet das inhaltliche Wissensfundament. Darüber hinaus existiert ein weiterer Wissenszugang über interaktive Probenbetrachter, die Rasterelektronenmikroskop-Aufnahmen visualisieren. Den dritten Wissenszugang bilden interaktive Formeln und 3D-Modelle von Gerätschaften wie dem Kontaktwinkelmessgerät. Lernende können hierbei z.B. über Schieberegler verschiedene Parameter anpassen und dazugehörige physikalische Effekte unmittelbar zuordnen (siehe Abbildung 1). Der Inhalt jedes Booklets, ist in sechs Unterabschnitte unterteilt – Orientierung, Theorie, Anwendung, Interaktion (mit Verlinkung zu Formeln und Proben), Reflexion und Quellen. Der implementierte Mechanismus zur Abfrage der Motivation hat bei dieser Variante keine Auswirkungen, die Ergebnisse dienen lediglich als Vergleichsmaß zu den adaptiven Versionen. Es handelt sich somit um die Kontrollgruppe der Untersuchung.

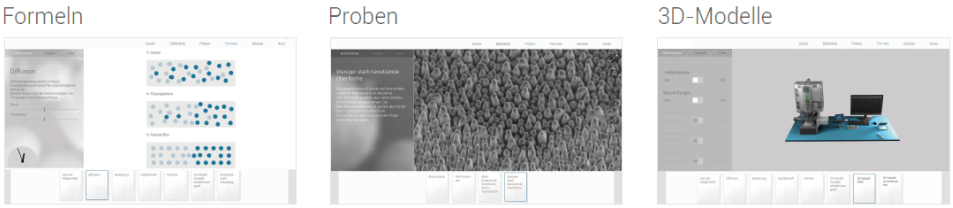



Abbildung 1: Interaktive Wissenszugänge der NanoTecLearn Plattform

### Direct Guidance

Die zweite Plattformversion ist eine adaptive Variante, welche die Direct Guidance-Technik in Form einer geführten Sequenzierung der Inhalte verwendet. Inhalt und Struktur der Booklets sind identisch zur nicht-adaptiven E-Learning Plattform. Hier hat die Motivationsabfrage jedoch einen Effekt, indem ein passendes nachfolgendes Unterkapitel vorgeschlagen wird. Je nach Einschätzung der aktuellen Motivation wird der Lernende zu unterschiedlichen Folgeabschnitten geführt, die Motivationseinbrüchen entgegenwirken sollen. Dem Lernenden wird dabei nach Abgabe der Bewertung eine Notifikation angezeigt: „Vielen Dank für die Einschätzung, es geht weiter mit Abschnitt X“. Dieses Feld ist mit dem entsprechenden Abschnitt verlinkt. Alle anderen, sonst frei zugänglichen Booklets und Unterabschnitte, sind bei Direct Guidance deaktiviert. Der Lernende muss in dieser Variante demnach dem Vorschlag folgen.

### Link Annotation

Die dritte Version verwendet die Link Annotation-Technik, die als Vorschlagssystem konzipiert ist. Auch hier werden die Selbsteinschätzungen wieder genutzt, um den motivational passfähigsten Folgeabschnitt zu ermitteln. Bei dieser Variante kann der Lernende jedoch selbst entscheiden, ob dem Vorschlag gefolgt werden soll oder nicht. Der empfohlene Folgeabschnitt wird in den Unterabschnitten farblich dezent hervorgehoben, ebenso erscheint der textuelle Hinweis: „Vielen Dank für die Einschätzung. Es wurden Empfehlungen für Sie hervorgehoben“. Die anderen Verlinkungen der Plattform sind hierbei nicht deaktiviert, sodass der Lernende sich frei in den Inhalten bewegen kann.

### Agent

Die vierte Version nutzt einen Agenten, der als Unterstützungssystem für den Lernenden gedacht ist, indem mithilfe einer Vorschlagsliste Anregungen für Folgeabschnitte gegeben werden. Der Agent ist in Form eines kreisförmigen Icons ständig im rechten oberen Rand des User Interface sichtbar und wird per Mausklick aktiviert. Nach der Motivationseinschätzung erscheint die Vorschlagsliste, z.B. mit Inhalten zum eigenen Ausprobieren in Form der interaktiven Proben oder Formeln (siehe Abbildung 2).

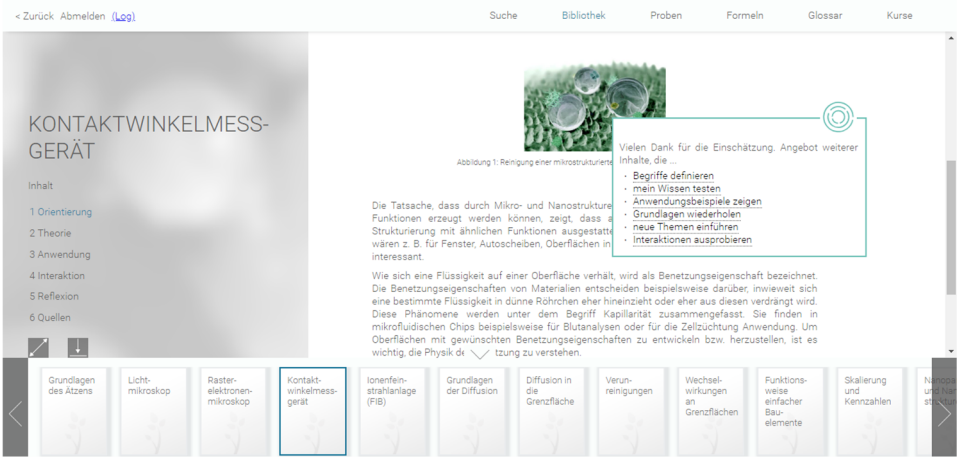



Abbildung 2: Vorschlagsliste der adaptiven NTL-Version Agent im Booklet Kontaktwinkelmessgerät

## 3.2 Motivations-Selbsteinschätzung

Am Ende jedes Abschnitts der Lernplattform soll der Lernende das aktuelle Motivationslevel mithilfe eines implementierten Self-Report-Mechanismus angeben. Dazu wurde je ein Item für die Parameter aktuell wirksames, situatives Interesse (In) und Erfolgszuversicht bzw. Selbstwirksamkeitserwartung (Ez) formuliert, das über die

Einschätzung „stimme nicht zu“ (-), „teils/teils“ (=), „stimme zu“ (+) erfragt wird. Grundlage hierfür bildete das Verständnis aktuell wirksamer Motivation nach Rheinberg und der entsprechende Fragebogen zur Erfassung der aktuellen Motivation sowie die verkürzte Version zur State-Messung der aktuellen Motivation [Vo03]. Besonders die State-Messung stellt das situativ wirksame Interesse und die Zuversicht, dem Lerninhalt gewachsen zu sein, in den Vordergrund. Dies wurde für die Formulierung eigener Items im konkreten Anwendungsfall wie folgt angepasst: „Die Inhalte des aktuellen Abschnitts haben mich interessiert“ und „Ich fühle mich den Inhalten und/oder Aufgaben gewachsen“. Der gebildete Motivationsindikator ist Basis für die Systemanpassungen der Benutzerführung. In der Lernplattform sind für die Unterabschnitte jedes Booklets alle Kombinationen der Motivations-Selbsteinschätzung in Form einer Tabelle hinterlegt sowie die zugehörigen Systemreaktionen. Diese wurden auf Basis einer Fokusgruppe mit Experten aus den Bereichen E-Learning-Didaktik, Content, Usability und technischer Umsetzung erarbeitet und evaluiert. Insbesondere bei (-)-Angaben werden möglichst motivationsförderliche Folgeinhalte angegeben. Bei ausbleibenden (-)-Angaben wird der Lernende seriell durch die Unterabschnitte des Booklets geführt.

### 3.3 Studiendesign

Aus den Ergebnissen der Fokusgruppen mit Studierenden ging hervor, dass das verspürte Kontrollempfinden, die Freiheitsgrade der Navigation und die allgemeinen Hilfestellungen des Systems besondere Bedeutung beigemessen wurde. Auf Basis dieser Erkenntnisse erfolgte im Sommersemester 2018 eine experimentelle Untersuchung der vier beschriebenen Plattformversionen mit 132 Studierenden der Technischen Universität Ilmenau. Für die Teilnahme war kein spezifisches Vorwissen nötig, daher erfolgte die Rekrutierung über den studentischen Mailverteiler der Universität. Die Teilnahme wurde mit einem Probandenhonorar vergütet. Einen Überblick über das Studiendesign mit den Messzeitpunkten der Untersuchung zeigt Tabelle 2.

| | **M1** | **M2** | **M3** |
|---|---|---|---|
| **Lerninhalt** | Vor der E-Learning Session | E-Learning Session in Form der Bearbeitung eines Kapitels der NanoTecLearn Plattform zum Thema Kontakt-winkelmessgerät | Nach der E-Learning Session |
| **Mess-instrument** | Fragebögen zu aktueller Motivation und Vorwissen | Logfiles zur Erfassung des Benutzerpfads durch das Lernmaterial und der beiden Motivations-indikatoren Interesse und Erfolgszuversicht | Fragebögen zu User Experience, Kontrollempfinden und Technologie-akzeptanz |

Tabelle 2: Überblick über Messgrößen und Messinstrumente zu den Messzeitpunkten M1-M3

Das Sample bestand aus 67 männlichen und 65 weiblichen Studierenden, die allesamt über 19 Jahre alt waren (M=24.42; SD=3,79). Die Probanden wurden randomisiert einer der vier Gruppen der experimentellen Untersuchung zugeordnet. Jede Version wurde final von 33 Studierenden getestet. Als Inhaltsbasis für die Studie wurde das Booklet zum Thema „Kontaktwinkelmessgerät" ausgewählt. Die Auswertung der Daten sowie die finale Überarbeitung und Festlegung auf eine adaptive Systemvariante wurden im Wintersemester 2018/19 vorgenommen. Die Ergebnisse der verwendeten standardisierten Fragebogen-Instrumente (vor und nach der E-Learning Session) sind zu finden in [Ba19]. Die nachfolgende Analyse fokussiert sich auf den eigentlichen Verlauf der Lernmotivation und die Wirksamkeit der Adaptation, also die konkrete E-Learning Session an sich.

### 3.4 Auswertung der Logfile-Analyse

Im direkten Vergleich mit den weiteren genutzten Fragebogeninstrumenten, die sich auf die Untersuchung der User Experience, des Kontrollempfindens, der aktuellen Motivation sowie der E-Learning Akzeptanz fokussieren, ermöglicht die Logfile-Analyse der E-Learning Sessions der Probanden zusätzliche Erkenntnisse. Zum einen ist der Motivationsverlauf während des Lernens mit NanoTecLearn nachvollziehbar, der über den Self-Report-Mechanismus erfasst wird. Zum anderen kann ermittelt werden, ob die auf Basis der Motivationseinschätzungen getroffenen Systemanpassungen motivational wirksam waren. Darüber hinaus lassen sich durch die erfasste Verweildauer Rückschlüsse auf das Themeninteresse ziehen. Im Fall der Link Annotation-Version konnte zudem ermittelt werden, wie häufig die Lernenden den Systemvorschlägen gefolgt sind, was Aussagen zur motivationalen Passfähigkeit ermöglicht. Die Logfiles wurden systemseitig erfasst und für jeden Probanden unter Angabe der anonymisierten Probandennummern gespeichert. Neben den Zeitangaben zur Verweildauer enthalten die Dateien Informationen über den aktuellen Abschnitt, in dem sich der Lernende befindet und über die getätigten Selbsteinschätzungen.

| **Unter-kapitel** | **26,1** | **26,2** | **26,3** | **26,4** | **26,5** | **26,6** |
|---|---|---|---|---|---|---|
| | Orientier-ung | Theorie | Anwendung | Interaktion | Reflexion | Quellen |
| **Bewertung Motivation** | **0** | **1** | **2** | **3** | **4** | **5** |
| | Ez -, In - | Ez =, In -<br>Ez -, In = | Ez +, In -<br>Ez -, In + | Ez =, In = | Ez +, In =<br>Ez =, In + | Ez +, In + |

Tabelle 3: Kodierung der Unterkapitel und Motivationseinschätzung

Die Booklets und Unterabschnitte wurden jeweils mit entsprechenden Ziffern codiert (dem in der Studie genutzten Booklet Kontaktwinkelmessgerät ist der Zahlencode 26 zugeordnet), ebenso die beiden Motivationsindikatoren und deren Ausprägung (siehe Tabelle 3). Die Bewertung der Motivation erfolgt jeweils als kombinierter Wert für beide Indikatoren (Ez und In) in Form einer fünfstufigen Skala, wobei ein Absinken der Motivation ((-)-Angabe für mindestens einen Indikator) jeweils schlechter bewertet wird als ein konstanter Wert (zweifache (=)-Angabe).

Nach Durchführung der Studie mussten die erhobenen Logfiles zunächst bereinigt und in eine auswertbare Form überführt werden. Der Fokus lag auf dem lernrelevanten Kapitel Kontaktwinkelmessgerät (26) und dessen Unterabschnitten. Die Exploration anderer Kapitel wurde für die Auswertung nicht erfasst. Bei Direct Guidance wurden außerdem die Besuche nach dem Unterkapitel Reflexion nicht mehr berücksichtigt, da die Version eine chronologische erste Bearbeitung vorsieht und diese nach der Reflexion beendet sein sollte. Ein wichtiger Bestandteil der Bereinigung war das Ausschließen von Abschnitten, die vorab definierte Kriterien nicht erfüllten. Dazu zählten Unterkapitel, die weniger als 10 Sekunden besucht wurden. Probanden, welche insgesamt weniger als 5 Minuten im Kapitel 26 verweilten, wurden vollständig ausgeschlossen, da hierbei nicht von einer ernsthaften Auseinandersetzung mit der Plattform ausgegangen werden konnte. Anschließend wurde die Reihenfolge des Durchschreitens der Unterkapitel für jeden Probanden in allen vier Versionen nachvollzogen. Darüber hinaus wurden die Motivations-Selbsteinschätzungen über die zugehörigen Zahlencodes (1-5) hinzugefügt.

# 4 Ergebnisse

Die nach dem Beenden eines Kapitels abzugebenden Einschätzungen zu Erfolgszuversicht (Ez) und Interesse (In) sind nachfolgend für jede der Versionen als gemeinsamer Motivationsparameter jeweils im Durchschnitt dargestellt sowie die durchschnittliche Verweildauer in den jeweiligen Abschnitten. Tabelle 4 gibt zunächst einen Überblick über die Anzahl der für die Auswertung zu verwendenden Datensätze und die Häufigkeiten der Einschätzungen pro Plattformversion.

| **Version** | **Link Annotation** | **Direct Guidance** | **Agent** | **Kontrollgruppe** |
|---|---|---|---|---|
| **Gültige Fälle (von n=33)** | 29 | 30 | 32 | 30 |
| **Häufigkeit der Einschätzung** | 132 | 114 | 115 | 174 |

Tabelle 4: Übersicht über die gültigen Fälle und Häufigkeiten der Einschätzung

Auffällig ist, dass alle Versionen im Unterkapitel 26,2, dem Theorieteil, einen erkennbaren Motivationseinbruch im Vergleich zum vorherigen Orientierungskapitel 26,1 aufweisen (vgl. Tabelle 5 und Abbildung 3). Dies ist ebenfalls der Visualisierung der Motivationsverläufe in den entsprechenden Abbildungen zu entnehmen. Es zeigt sich zudem, dass die interaktiven Inhalte die Motivation wieder merklich steigern konnten, was mit den Interviewaussagen einer ersten Studie zur Lernmotivation im Wintersemester 2017/18 mit der nicht-adaptiven NanoTecLearn Plattform übereinstimmt.
Die Motivation, in der sich dem interaktiven Abschnitt anschließenden Reflexion, bleibt relativ unverändert, jedoch war über die Freitextantworten zum Ende der Befragung zu erkennen, dass die Probanden sich in diesem Abschnitt ebenfalls mehr interaktive Elemente, z.B. in Form von interaktiven Wissenstests gewünscht hätten.

| Version | Link Annotation | Direct Guidance | Agent | Kontrollgruppe |
|---|---|---|---|---|
| **Durchschnitt Motivations-einschätzung pro Abschnitt** | 26,1 – 3,77<br>26,2 – 2,96<br>26,3 – 2,96<br>26,4 – 3,70<br>26,5 – 3,92<br>26,6 – 3,00 | 26,1 – 4,42<br>26,2 – 3,42<br>26,3 – 3,55<br>26,4 – 4,30<br>26,5 – 4,33<br>26,6 – k.A. | 26,1 – 4,07<br>26,2 – 3,55<br>26,3 – 3,92<br>26,4 – 4,64<br>26,5 – 3,95<br>26,6 – 2,80 | 26,1 – 4,19<br>26,2 – 2,93<br>26,3 – 2,68<br>26,4 – 4,08<br>26,5 – 4,04<br>26,6 – 2,63 |
| **Durchschnitt Verweildauer pro Abschnitt** | 26,1 – 01:50<br>26,2 – 04:54<br>26,3 – 04:30<br>26,4 – 03:34<br>26,5 – 02:31<br>26,6 – 01:37 | 26,1 – 02:18<br>26,2 – 05:58<br>26,3 – 05:31<br>26,4 – 04:41<br>26,5 – 02:16<br>26,6 – k.A. | 26,1 – 01:46<br>26,2 – 04:26<br>26,3 – 03:32<br>26,4 – 02:46<br>26,5 – 02:06<br>26,6 – 00:59 | 26,1 – 02:13<br>26,2 – 05:24<br>26,3 – 03:39<br>26,4 – 03:10<br>26,5 – 01:26<br>26,6 – 01:08 |

Tabelle 5: Vergleich der durchschnittlichen Motivationseinschätzungen und Verweildauer (Notation pro Plattformversion: Unterkapitel – Mittelwert)

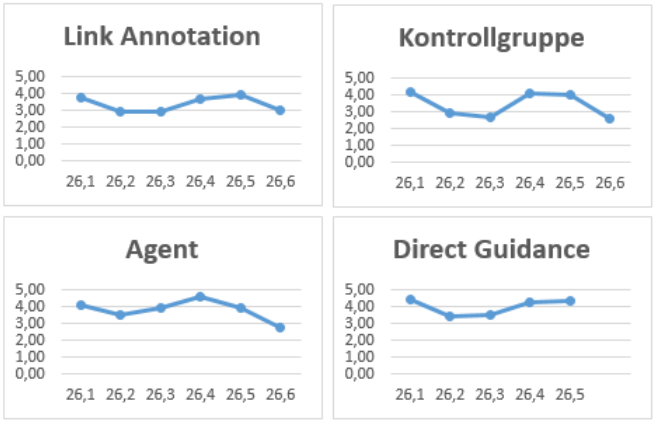



Abbildung 3: Vergleich der Motivationsverläufe der unterschiedlichen Plattformversionen

Die Daten aus Tabelle 5 wurden zudem einer einfaktoriellen Varianzanalyse (ANOVA) unterzogen, um festzustellen, ob sich die Mittelwerte der Motivationseinschätzung pro Abschnitt in den Versuchsgruppen unterscheiden. Das Ergebnis des F-Tests zeigte, dass sich die Motivationseinschätzungen in den Versuchsgruppen nicht signifikant unterscheiden, $F(3, 20)=0{,}297$, $p=0{,}827$. Eine weitere ANOVA sollte zeigen, ob sich die Mittelwerte der beiden Motivationsparameter in den unterschiedlichen Lernabschnitten der Plattformversionen unterscheiden. Für Erfolgszuversicht konnte kein signifikanter Unterschied festgestellt werden. Es ergab sich jedoch ein signifikanter Unterschied für Interesse zwischen den Versuchsgruppen für die Lernabschnitte Orientierung [$F(3, 117) = 2{,}691$, $p=0{,}049$], Theorie [$F(3, 112) = 3{,}187$, $p = 0{,}027$], Anwendung [$F(3, 106) = 5{,}872$, $p = 0{,}001$] und Reflexion [$F(3, 87) = 4{,}552$, $p = 0{,}005$]. Als Post-Hoc-Test wurde

anschließend der LSD-Test berechnet. Das Ergebnis der Mehrfachvergleiche zeigte, dass im Lernabschnitt Orientierung Unterschiede im Interesse nur für Direct Guidance und Link Annotation sowie die Version Agent auf dem Niveau 0,05 signifikant sind. Innerhalb des Lernabschnitts Theorie sind nur zwischen den Versuchsgruppen Link Annotation und der Kontrollgruppe sowie zwischen Direct Guidance und der Version Agent keine Mittelwertunterschiede für das Interesse zu verzeichnen. Innerhalb der Anwendung ist die Differenz der Interessen-Mittelwerte für die Kontrollgruppe und Direct Guidance sowie die Agentenversion auf dem Niveau 0,05 signifikant und ebenso für die Gruppen Link Annotation und Agent. Für den Lernabschnitt Reflexion unterscheidet sich das Interesse nur für die Kontrollgruppe signifikant von allen anderen Gruppen, $p < 0,05$.

Die Auswertung der Link Annotation-Version ermöglicht zudem die Interpretation der Anzahl der befolgten Vorschläge. Von 132 Vorschlägen wurden 86 befolgt, dies entspricht einer Quote von 65%, was zeigt, dass eine Mehrzahl der Vorschläge als motivationsförderlich empfunden wurde. Die Nicht-Befolgung kann aus der fehlenden inhaltlichen Passfähigkeit oder dem Übersehen der visuellen Hinweise resultieren. Bei vielen Probanden ist die Motivation während der Bearbeitung kaum merklich gesunken. Über offene Fragestellungen zum Ende der Erhebung und auf Basis der Ergebnisse eines User Experience- und Technologieakzeptanz-Fragebogens [Ba19] konnte ermittelt werden, dass viele Studierende die bloße Existenz einer solchen E-Learning Plattform bereits als motivierende Ergänzung zu klassischen Lehrmaterialien wahrgenommen haben. Probanden mit wenig domänenspezifischem Vorwissen, gaben darüber hinaus an, dass die geführte Sequenzierung der Direct Guidance als hilfreich und motivierend empfunden wurde. So ist in Abbildung 1 und Tabelle 5 ersichtlich, dass in dieser Version die geringsten Motivationsschwankungen und gleichzeitig die über alle Unterkapitel höchsten Mittelwerte der Motivationseinschätzung auftreten.

Kritisch anzumerken ist, dass die selbstregulierte Lernsituation und das experimentelle Setting an sich bereits Einfluss auf die Aussagekraft der Ergebnisse genommen haben. So wurden einerseits Einschätzungen teilweise nicht vollzogen, da zum Beispiel das Kapitel nicht komplett bearbeitet wurde und die am Ende der Kapitel implementierte Selbsteinschätzung schlicht übersehen wurde. Andererseits konnten Einschätzungen doppelt vollzogen werden, wenn zwischenzeitlich das Booklet verlassen wurde. Auch konnten aufgrund der separaten Darstellung der Proben und Formeln, gleichzeitig Proben anderer Booklets betrachtet werden. Dies kann sowohl positive als auch negative Effekte auf die Motivation haben, da beispielsweise das Interesse auf andere Themen steigen könnte oder aber die Motivation aufgrund der Orientierungslosigkeit sinkt. Hier stößt die Aussagekraft der formalisierten Motivationseinschätzung an ihre Grenzen.

Die einzelnen Adaptationstechniken bilden zusätzliche individuelle Einflussfaktoren. So konnten sich die Probanden bei der Direct Guidance-Version nach erstmaligem Beenden des Booklets frei auf der Plattform bewegen. Die anschließende freie Exploration wurde in der Analyse nicht berücksichtigt, um eine Vergleichbarkeit mit den anderen Versionen gewährleisten zu können. Im Fall der Link Annotation-Version wurden Hervorhebungen der empfohlenen Kapitel und bei der Agenten-Version das Symbol des Agenten öfter nicht wahrgenommen (in diesem Fall konnten auch keine Einschätzungen vorgenommen werden).

## 5 Ausblick und Diskussion

Die Ergebnisse der Logfile Analyse in Kombination mit den fragebogenbasierten Erhebungen sowie offenen Fragestellungen im Anschluss an die E-Learning Session münden in folgenden Empfehlungen, die als Ausgangspunkt für die Gestaltung motivationsförderlicher adaptiver E-Learning Plattformen dienen können:

- Adaptationen sollten visuell und textuell angemessen codiert werden, um vom Lernenden eindeutig identifiziert werden zu können
- Geführte Sequenzierung durch E-Learning Inhalte eignet sich besonders für Lernende mit wenig domänenspezifischem Vorwissen
- Um Motivationseinbrüchen entgegenzuwirken, sollte eine ausgewogene Aufteilung zwischen Textpassagen und interaktiven Inhalten vorgenommen werden
- Interaktive Inhalte und deren angemessene Einbindung wirken nach längeren Textabschnitten besonders motivationsfördernd
- Wissensabfragen und Zusammenfassungen sollten möglichst interaktiv sein und direktes Feedback geben
- Heterogenen Zielgruppen profitieren vom Angebot wählbarer Lernziele und unterschiedlicher Anspruchsniveaus

Auf Basis dieser Erkenntnisse wurde die Lernplattform für eine nachfolgende Studie optimiert. Der finale motivationsbasierte Adaptationsmechanismus wird ebenfalls Gegenstand einer Laborevaluation mit Studierenden sein. Es soll somit eine Abschätzung ermöglicht werden, ob der zusätzliche Entwicklungsaufwand gegenüber einem nicht-adaptiven E-Learning System für Lehrende gerechtfertigt ist. Zukünftige Untersuchungen sollten sich zudem einer umfassenderen Erhebung der Lernmotivation während des Lernens mit einer adaptiven E-Learning Session widmen. So wären z.B. Langzeitstudien mit der Experience Sampling Methode denkbar, um den Motivationsverlauf über den Lernprozess eines Semesters darzustellen.

## Dank

Das Projektvorhaben SensoMot wird gefördert durch das Bundesministerium für Bildung und Forschung (BMBF) unter der Fördernummer 16SV7516 im Programm “Erfahrbares Lernen”.

Literaturverzeichnis

[Ba18] Bauer, M.; Bräuer, C.; Schuldt, J.; Krömker, H.: Adaptive E-Learning for Supporting Motivation in the Context of Engineering Science. In (Nazir, S., Teperi, A.-M., Polak-Sopinska, A., Hrsg.): Advances in Human Factors in Training, Education and Learning Sciences, Proceedings of the AHFE 2018 International Conference on Human Factors

in Training, Education and Learning Sciences, Advances in Intelligent Systems and Computing, Vol. 785, Orlando, S. 409-422, 2018.

[Ba19] Bauer, M.; Bräuer, C.; Schuldt, J.; Krömker, H.: Evaluation of the User Experience of Adaptation Techniques to foster Learning Motivation in an adaptive E-Learning Platform. In: INTED2019, Proceedings of the 13th annual International Technology, Education and Development Conference, Valencia, 2019.

[Br07] Brusilovsky, P.: Adaptive Navigation Support. In (Brusilovsky, P., Kobsa, A., Nejdl, W., Hrsg.): The Adaptive Web: Methods and Strategies of Web Personalization. Springer, Berlin, Heidelberg, S. 263-290, 2007.

[Gr14] Graf, S; Kinshuk: Adaptive Technologies. In (Spector, J. M., Merrill, M. D., Elen, J., Bishop, M. J., Hrsg.): Handbook of Research on Educational Communications and Technology. Springer, New York, S. 771–779, 2014.

[Kn09] Knutov, E.; de Bra, P.; Pechenizkiy, M.: AH 12 years later. A comprehensive survey of adaptive hypermedia methods and techniques. In: New Review of Hypermedia and Multimedia, 15/01, S. 5-38, 2009.

[Kr16] Krömker, H.; Hoffmann, M.; Huntemann, N.: Wissensstrukturierung für das Lernen in den Ingenieurwissenschaften. In (Kammasch, G., Klaffke, H., Knutzen, S., Hrsg.): Technische Bildung im Spannungsfeld zwischen beruflicher und akademischer Bildung. Die Vielfalt der Wege zu technischer Bildung: Referate der 11. Ingenieurpädagogischen Regionaltagung 2016, Berlin, S. 101–108, 2016.

[Le11] Leutner, D.: Adaptivität und Adaptierbarkeit beim Online-Lernen. In (Klimsa, P., Issing, L., Hrsg.): Online-Lernen. Planung, Realisation, Anwendung und Evaluation von Lehr- und Lernprozessen online. De Gruyter, München, S. 115–123, 2011.

[Pa10] Paramythis, A.; Weibelzahl, S.; Masthoff, J.: Layered Evaluation of Interactive Adaptive Systems. Framework and Formative Methods. In: User Modeling and User-Adapted Interaction 20/05, S. 383–453, 2010.

[Sc17] Schneider, O.; Martens, T.; Bauer, M.; Ott-Kroner, A.; Dick, U.; Dorochevsky, M.: SensoMot – Sensorische Erfassung von Motivationsindikatoren zur Steuerung adaptiver Lerninhalte. In (Igel, C., Ullrich, C., Wessner, M., Hrsg.): Bildungsräume 2017: Gesellschaft für Informatik, Bonn, S. 267–272, 2017.

[Sh12] Shute, V. J.; Zapata-Rivera, D.: Adaptive Educational Systems. In (Durlach, P. J., Lesgold, A. M., Hrsg.): Adaptive Technologies for Training and Education. Cambridge University Press, Cambridge, S. 7-27, 2012.

[So15] Sosnovsky, S.; Brusilovsky, P.: Evaluation of Topic-based Adaptation and Student Modeling in QuizGuide. In: User Modeling and User-Adapted Interaction, 25/04, S. 371-424, 2015.

[Vo03] Vollmeyer, R.; Rheinberg, F. (2003): Aktuelle Motivation und Motivation im Lernverlauf. In (Stiensmeier-Pelster, J., Rheinberg, F., Hrsg.): Diagnostik von Motivation und Selbstkonzept. Göttingen, Bern: Hogrefe, S. 281-296.

# "What can I help you with today?"

## Exploring Opportunities of Learner Modeling for Online Educational Portals

Sebastian Wollny[1], Jan Schneider[1], Marc Rittberger[1] and Hendrik Drachsler[123]

**Abstract:** Online educational Portals (OEPs) subsume a field of online repositories for a wide range of stakeholders. They are characterized by easily accessible structures that do not require visitor accounts. Most OEPs therefore provide content in one way to all visitors also known as "one size fits all" approach. With this study, we examine if Web Analytics data can be used to infer the modeling of learners for OEPs. This would be the basis for additional and more personalized ways of providing content to various stakeholders. In order to draw conclusions about opportunities and limitations of Web Analytics in this regard, the data structure of the Fachportal Pädagogik, as one of the largest educational OEPs in Germany, is compared with a common Learner Modeling Framework. The evaluation of the results finally leads to two major challenges that must be overcome in order to achieve personalized content and learning experiences on OEPs.

**Keywords:** Learner Modeling, Personalization, Adaptive Hypermedia, Learning Analytics, Recommender Systems, Open Educational Resources, Online Libraries, Online Educational Portals, Digital Assistants, Chatbots

## 1. Introduction

Online educational portals (OEP) are a container term for a variety of repositories containing online learning and teaching materials. The diversity of OEP repositories ranges from well-established online libraries over learning support services, such as edutags.de, until OER portals such as Merlot.org or the Education Portal and educational content servers such as Bildungserver.de in Germany. The Open Data movement is currently also creating new types of OEPs that will deliver content and data sets to a wide audience.

[1] DIPF | Leibniz Institute for Research and Information in Education, Educational Technologies, Rostocker Straße 6, D-60323 Frankfurt am Main, {wollny | schneider.jan | rittberger | drachsler}@dipf.de
[2] Goethe-Universität Frankfurt am Main, Educational Technologies, Robert-Mayer-Str. 10, D-60054 Frankfurt am Main [lastname@em.uni-frankfurt.de]
[3] Open University of the Netherlands, Welten Institute, Valkenburgerweg 177, NL-6419 AT Heerlen, [firstname.lastname@ou.nl]

doi:10.18420/delfi2019_348

Independent from the kind of data hosted in OEPs, they are all characterized by an easily accessible structure that can facilitate services for a variety of stakeholders at any moment. They are designed to fulfill the demands of various stakeholders (learners, teachers, curators, editors, etc.) [CEP14], [Ch18]. However, visitor accounts proved to be impracticable in the past as they were not used intensively. As a result, OEPs restricted themselves to a "one size fits all" approach that treats all visitors uniformly. Studies have shown a drop of visitors attributing it to the “one-size-fits-all” approach [CL07], [Jo09], [Wi07]. This raises the question of the "one size fits all" approach is still up-to-date, or whether personalized content is better suited to individual information needs.

In this paper we analyze the web analytics data structure obtained from the Fachportal Pädagogik, one of the largest scientific OEPs in Germany, to examine the opportunities and technical limitations of Web Analytics for modeling learners in the context of an OEP.

Based on the findings of the Web Analytics investigation, we argue that educational portals need to focus more on learner modeling in order to overcome the one-size-fits-all approach and increase their usability, educational effectiveness and general level of satisfaction [AB08]. Especially with the upcoming open data repositories[4] joining the OEP family, it would be a great advantage, if OEPs would support the diversity of stakeholders with personalized content for individual information needs and develop beyond the one-size-fits-all approach.

The structure of this paper is as follows, first we shortly describe the state of the art of *Learner Modeling* and how it could be applied for OEPs (section 2), thereafter we review current information that can be taken out of Web Analytics in OEPs (section 3). Next, we introduce our Web Analytics study (section 4) and present the main findings (section 5). Finally, we discuss our results (section 6) and conclude our research with two challenges that OEPs have to overcome to fulfill the information needs of their stakeholders in the 21st century.

## 2. State of the Art on Learner Modeling

Looking at *Learner Modeling*, it is a technique that is strongly linked to the personalization of learning or adaptive learning [CV13]. Since most stakeholders of OEPs have specific information needs, we consider them as learners. The term *Learner Modeling* traditionally refers to the modeling of learners related to knowledge diagnosis and adaptive scaffolding. Through intensive research on User Modeling [BM07], the scope of this traditional *Learning Modeling* approach is extended and therefore used in

[4] An overview for Germany is available on: https://open-educational-resources.de , 24.03.2019

this form. Especially in OEPs, adaptations or personalization of content could help learners to orient and structure their information needs [BR09]. Adaptations to Learners are in many OEP cases limited to recommendation systems, which rely on collaborative or semantic filtering techniques [Ve12], [Dr15] and are therefore often driven by a content-based approach.

*Learner Modeling* can be divided into three phases according to [Ga07]: 1. Data Collection, 2. Profile Construction and 3. Profile Application (Fig.1).

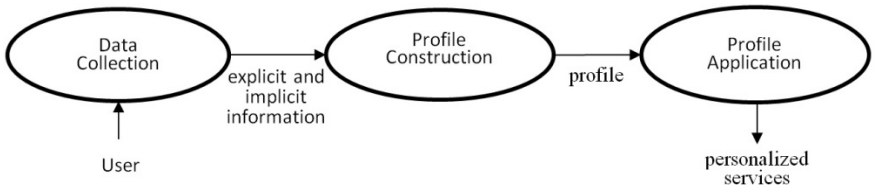



Fig. 1: Learner Modeling process [Ga07]

The data collected in the first phase can be distinguished between explicit and implicit. Among other behavioral data, the browser cache, search logs, and especially weblogs can be used to determine implicit learner profile data. In the next phase, collected data can be transferred to a learner profile, which is then applied in the third phase. For providing personalization based on *Learner Models*, it can be used for memorization, guidance, customization and task performance support of learners [Ca09].

As our study mainly focuses on creating learner models (profiles) out of collected Web Analytics data, we especially focus on phase one and two of the modeling process. A prerequisite for an extension of adaptations in the context of OEPs is the correct extraction of learner characteristics (Data Collection), which can be combined to a learner model (Profile Construction). There are several models of learners presented in previous work [RL08], [ND08], [Sp00]. For this paper, we selected a *Learner Modeling Framework* that is based on Brusilovsky and Millán [BM07]. This *Learner Modeling Framework* seems to be most appropriated for our purpose, as it is a common model in Learning Analytics [CMS17], Adaptive Learning [MDK11] and Recommender Systems for Learning [Ve12] to describe the characteristics of learners that are supported with personalized content. It consists of six characteristic categories about learners, which are put together to an overall model of a learner. These categories are 1. Knowledge, 2. Interests, 3. Goals and Tasks, 4. Background, 5. Individual Traits, 6. Context of Work. In the following, the categories are presented and explained by examples.

**Knowledge**

A learner's knowledge is technically often represented by directed graphs or ontologies. It is a changeable feature, as learners are able to learn and forget. Knowledge features are often considered important because they reflect the level of understanding in relation to the domain. After extraction, they can be used to adapt the learning process in the way, that it helps to expand knowledge and avoids unwanted repetitions.

**Interests**

Interest reflects the attention or focus of learners. It can be grouped into subject areas present in the respective domain and described by weighted keyword vectors or overlay models. Learner interests are an important reference point for generating learning paths. Especially for use cases in which certain subject areas are examined, current interests define directions for further exploration. This type of learner feature is often used as a core component of recommender systems to support learner exploration of repositories.

**Goals and Tasks**

Learning goals and tasks are related features. Goals are personally desired states of learners in the future. They are the cause of personal tasks that have to be mastered in order to reach the goal. Goals and tasks reflect the information needs of learners. They reflect the information needs of learners and can be selected or defined in adaptive learning systems. The nature of personal learning goals incorporates the relation to different time periods. While graduating from school can be a long-term goal, short-term goals are, for example, finding the solution of an equation.

**Background**

The learning background is a feature category that shares similarities with knowledge features. It defines past experiences that are not directly related to the respective domain. The background features of learners can be combined into personas or stereotypes. While a professor has a larger research background, a student has a more curricular background.

**Individual Traits**

Individual character traits are personal characteristics that cannot be classified in any of the previous categories. These are mainly preferences and habits that affect the learning process. These kinds of features can be extracted through psychological tests.

**Learner Context**

The learner context consists of all the features that describe the learner's current situation. In addition to the use of device types, the learner context also includes the learner's location, the affective state, social and personal context, and physical environment.

## 2.1. Example of applying a Learner Model to an OEP

The following table (Tab. 1) shows two examples of learners using an OEP. Both learners would like to find out something about self-regulated learning (SRL) but have different reasons and strategies. While Anna prefers to look for a variety of practical content on SRL, Bob would prefer a single meta-study. In addition, Anna would be happy about a paper in the context of biology teaching, while Bob prefers to look for

general insights. Both have different features in all feature categories, which could also lead to a differentiated treatment of these learners.

| Visitor | Knowledge | Interests | Goals | Tasks | Back-ground | Indiv. Traits | Work Context |
|---|---|---|---|---|---|---|---|
| Anna | Didactics, Biology | Teaching Methods | Biology Lesson Preparation | Find SRL Research | Teacher | Wants to understand things in detail | Last year students wanted to work independently, Desktop, Evening |
| Bob | Psychometrics, Statistics | Mathematics, Psychology | Finding evidence for SRL Model | Analyze State-of-the-Art Papers | Researcher | Likes fast + precise results | Preparation of a Presentation, at work |

Tab.1: Learner Model Examples

## 3. Web Analytics

Collecting data about learners is a requirement for modeling them. Web Analytics data can be seen as an indication for processes on websites. Similar to key performance indicators (KPI) from the business sector [CP12], these Web Analytics indicators are important for the interpretation of visitor actions on websites [Ke11]. Also for learning on OEPs Web Analytics data could be suitable to personalize learning. To support *Learner Modeling* on OEPs, Web Analytics measuring methods have the advantage of being universally applicable to a wide range of OEPs and common for data collection and analysis on these portals.

Web Analytics is defined as the collection, measurement, analysis, and reporting of web data with the main purposes to identify web traffic and usage patterns [BBC07]. The data usually comes from four sources: 1. Direct HTTP request data, 2. Network level and server-generated data associated with HTTP requests, 3. Application level data sent with HTTP requests, and 4. external data [ZP15].

In contrast to local device recordings, which provide a detailed learner model, website recordings have the advantage to collect learner data of large groups without the requirement to manually install software on each single device [Ga07].

In the following, we group the major Web Analytics techniques into the following four categories that we will also use of reviewing the Web Analytics results: A. Click-Stream Analysis, B. Session Analysis, C. Visitor Analysis, and D. Event Analysis.

**A. Click Stream Analysis**. The most often used methods can be subsumed by the Click-Stream Analysis. It contains the logging of click and HTTP requests that can be assembled to a personal usage log of a website. These clicks and requests are grouped into sessions or visits, which are defined as active time periods of visitors on the website. As indicators for the attractiveness of pages the click-through rate, bounce

rate and the time on page can be used. They express if the visitor stays on the website or rather browses to another website.

**B. Session Analysis**. For the analysis of sessions, four data types are important: 1. The landing page, which is the first visited page of a visitor, 2. The exit page, which is the last visited page of a visit, 3. The visit duration and 4. The referrer. Latter can be seen as the external page, which linked to the website and caused the landing page request. Nowadays a large amount of website traffic is referred by search engines.

**C. Visitor Analysis.** The visitor itself can be analyzed by the information they share with the website through their browsers. This could be browser settings or cookies, which are also used to identify the visitor. Whether a visitor is new or returning can be classified in this way too. An identification method, called Fingerprinting, is based on the measurement of subtle differences in communication between browsers and servers, which are caused by these individual configurations. This measurement and the associated analysis is carried out on the server side, which is why large visitor groups can be covered with it.

**D. Event-based Analysis.** These events are specific to the purpose of a website. For commercial websites, an event could be considered for example as a click on the "buy" button. Events can be defined flexibly. They enable website owners to protocol all necessary and interesting actions. By analyzing website events, the effectiveness of online advertisement campaigns or the effects of website adaptions can be measured.

With these four major categories in mind, we will examine in section 4, how data structures from Web Analytics can be matched to the Learner Modeling Framework.

## 4. Method

In this paper, we explore whether the use of Web Analytics techniques on the log data of educational portals can be used to create a learner model for an OEP without having to resort to visitor accounts. Our main Research Question is, therefore:

**RQ1: Can we use Web Analytics log data of OEPs in order to create a learner model based on the Learner Modeling Framework?**

In order to identify to what extent it is possible to create a model of learners, based on the *Learning Modeling Framework*, we guide our examination with the use of the following sub-questions:

**RQ1a: Can we use Web Analytics to identify learner knowledge?**

**RQ1b: Can we use Web Analytics to identify learner interest?**

**RQ1c: Can we use Web Analytics to identify learner goals and tasks?**

**RQ1d: Can we use Web Analytics to identify learner backgrounds?**

**RQ1e: Can we use Web Analytics to identify learner individual traits?**

**RQ1f: Can we use Web Analytics to identify learner context?**

To give answers to these questions we will analyze the Web Analytics log data structure of the Fachportal Pädagogik[5]. The findings to sub-questions are rated on two five-point scales, which we call the Applicability Score and Variety Sore. While the Applicability Score describes the quality of data has been found, the Variety Score describes the number of deducible features has been found. The scales of the scores are described by the following table (Tab. 2):

| **Applicability Score** | | **Variety Score** | |
|---|---|---|---|
| 1 | No applicable data found (but indications may be possible) | 1 | No feature values deducible |
| 2 | Weak indications found | 2 | Small number of feature values deducible |
| 3 | Strong indications found | 3 | Medium number of feature values deducible |
| 4 | Noisy feature values found | 4 | Large number of feature values deducible |
| 5 | Applicable feature values found | 5 | Wide variety of feature values deducible |

Tab. 2: Applicability Score scale and Variety Score scale

### 4.1. Data Source – Fachportal Pädagogik

The Fachportal Pädagogik is one of the most visited OEPs in the context of scientific education in Germany. With over 800.000 Online-Sessions per year, it offers its visitors access to an overview in digital databases, scientific full-texts and library registries. The data pool includes more than 1.000.000 entries on educational literature and obtains its content primarily from library networks.

We used the Matomo[6] Web Analytics engine in order to track and analyze the data of the Fachportal Pädagogik. It enables state-of-the-art Web Analytics with the possibility to access raw data. Data obtained in this way is mainly stored in three tables. While the log action table contains all actions that have been detected so far, the log visit table is a

[5] Reachable through https://www.fachportal-paedagogik.de/ , 24.03.2019
[6] Until December 2017 Matomo was called "Piwik".

record of session data. Actions can be mainly seen as visited URLs, which are assigned to a session by the "log link visit action" table.

By examining the elements of the Web Analytics log data tables we can identify the information types of the four Web Analytics technique categories (A to D) in Tab. 3. It indicates which of the three Matomo data tables are required to extract related information types. These must be considered in many cases in the context of the portal's content and have to be combined with it to draw conclusions. In our case, this connection is made by the URL in the 'log action' table, since it determines the page content.

| Information type | log action | log visit | log link visit action |
|---|---|---|---|
| **A. Click-Stream Analysis** | | | |
| *Click Paths* | **x** | | **x** |
| *Time spent on pages* | **x** | **x** | **x** |
| *Classification of requested pages* | **x** | **x** | **x** |
| **B. Session Analysis** | | | |
| *Session indicators* (visit duration, referrer, returning visitor, …) | | **x** | |
| *Session behavior* (e.g. exploration / exploitation) | **x** | | **x** |
| *Topic- and content type-wise classification of sessions* | **x** | **x** | **x** |
| **C. Visitor Analysis** | | | |
| *Visitor identification* | | **x** | |
| *Time since the last visit* | | **x** | |
| *Visitor preferred language* | | **x** | |
| *IP address* | | **x** | |
| *Visitor location* (incl. time zone) | | **x** | |
| *Device information* (type / brand / OS version / screen resolution, …) | | **x** | |

| | | | |
|---|---|---|---|
| *Browser information* ( type / version / plugins ) | | **x** | |
| **D. Event-based Analysis** | | | |
| *Bounce rate* | **x** | **x** | **x** |
| *Click-through rate* | **x** | **x** | **x** |

Tab. 3: Overview of Web Analytics information types in Matomo log data.

## 5. Results

In the following, we present the findings we derived by comparing Web Analytics data and exemplary data table contents retrieved by a perennial record with feature categories of the *Learner Modeling Framework* in order to identify technical opportunities and limitations of modeling learners on OEPs. In addition, we rate the opportunity of Web Analytics to learner modeling with the Applicability Score (AS) and the Variety Score (VS) for each aspect of the *Learner Model Framework*.

**RQ1a - Knowledge**: Knowledge is usually inferred through assessments. We were not able to identify any type of knowledge assessment in the OEPs. However, we could identify some week indicators about the learner's knowledge. For example, we could assume that after visiting a page, the learner has some knowledge about its content. **AS: 2, VS: 2**

**RQ1b - Interests:** The learner's click-path data informs us about the learner's attention, which in turn provides some indications about the learner's interest. The available data is limited to the interaction with the portal, therefore the identified interest of the learner is restricted to these interactions. **AS: 5, VS: 4**

**RQ1c - Goals and Tasks:** The OEP has no feature where it asks the learner directly for their current goals or tasks. Therefore, we were not able to identify any strong indicators about them. Nonetheless, we argue that for long sessions click-path can be used as an indicator of the current task. **AS: 2, VS: 3**

**RQ1d - Background:** The strongest indicator identified for inferring the background of the learner is the access to content that was created specifically for a visitor group (e.g. content for elementary school grammar teachers). However, not all the content stored in the portals is targeted only for specific groups. Prolonged tracking of the learner could strengthen this indicator by looking at more content accessed by the learner, as well as identifying their level of familiarity with the portal. **AS: 2, VS: 2**

**RQ1e - Individual Traits:** Web usage behavior is the only indicator we have been able to identify that is related to the individual characteristics of the learner. This provides us only a week indication because the generalization of visitor behavior to individual

character traits is a complex process. Moreover, the possibilities to adapt the OEP in our case on the basis of specific individual traits are very limited. **AS: 3, VS: 2**

**RQ1f - Learner Context**: Regarding the learner context we could identify the location, local time, device being used by the learner. The practical personalization use of this type of contextual data is limited. Important aspects of the learner context such as affective state, social and personal context, and the physical environment were not identified. **AS: 2, VS: 2**

## 6. Discussion

The aim of this research is to assess the technical possibilities of *Learner Modeling* through Web Analytics. We retrieved results in all sub-questions and want to use them to answer the main Research Question RQ1 (Tab. 4):

| | **Knowledge** RQ1a | **Interests** RQ1b | **Goals and Tasks** RQ1c | **Background** RQ1d | **Indiv. Traits** RQ1e | **Work Context** RQ1f | **Ø** |
|---|---|---|---|---|---|---|---|
| AS | 2 | 5 | 2 | 2 | 3 | 2 | 2,7 |
| VS | 2 | 4 | 3 | 2 | 2 | 2 | 2,5 |

Tab. 4: Overview of Application Score and Variety Score ratings

The findings show limitations of the Web Analytics data structure in all feature categories. Besides RQ1b (Identification of learner interests), the data structure of Web Analytics engines does not generally allow to answer the sub-questions (RQ1a,c,d,e,f) positively. In terms of modeling knowledge (RQ1a), goals or tasks (RQ1c), the learner's background (RQ1d), individual traits (RQ1e) and learner's context (RQ1f) we found that there is the possibility that Web Analytics could give indications for some sessions or learners.

This indicates that data obtained through Web Analytics is already significantly limited in terms of applicability (Ø = 2,7) and variety of deducible learner features (Ø = 2,5) by its structure to model learners. Although Web Analytics seems to be suitable to model the interest of learners (RQ1b), many applications in this context are already covered by recommender systems. In addition, the interest feature can change dynamically over time and with context. A longer observation period is likely to improve data quality but not expected to lead to fundamental changes in the assessment.

In order to overcome the "one size fits all" approach of OEPs, alternative tools should be considered in addition to Web Analytics. In this context, interactive assistants in the form of Chatbots, new forms of search engines or content presentations are conceivable. Although learner profiles retrieved through Web Analytics foreseeable are too limited to represent the learner as a whole, they reflect the individual interaction of a resource and

must be considered in the context of data privacy. This raises two major challenges that need to be addressed in order to fulfill the OEPs stakeholder's information needs in the 21st century:

The first challenge can be named as "measurement problem". It is concerned about the question, how we can obtain missing data that cannot be found in the current OEP website interaction record. The second challenge can be named as "focus problem". It is concerned about the question, which learner features are needed for which OEPs in order to enable personalized learning experiences.

Further exploration of these challenges would help to understand personalized learning for OEPs, open education and educational online services.

Bibliography

[AB08] Arh, T.; Blazic, B. J.: A case study of usability testing–the SUMI evaluation approach of the EducaNext portal. In WSEAS transactions on information science and applications, 2008, 5; pp. 175–181.

[BBC07] Burby, J.; Brown, A.; Committee, W. S.A.: Web analytics definitions. In Washington DC: Web Analytics Association, 2007.

[BM07] Brusilovsky, P.; Millán, E.: User models for adaptive hypermedia and adaptive educational systems: The adaptive web. Springer, 2007; pp. 3–53.

[BR09] Blees, I.; Rittberger, M.: Web 2.0 learning environment. Concept, implementation, evaluation. PAU Education, 2009.

[Ca09] Castellano, G. et al.: Innovations in web personalization: Web Personalization in Intelligent Environments. Springer, 2009; pp. 1–26.

[CEP14] Camilleri, A. F.; Ehlers, U. D.; Pawlowski, J.: State of the art review of quality issues related to open educational resources (OER). Luxembourg: Publications Office of the European Union, 2014.

[Ch18] Chounta, I.-A.: A review of the state-of-art of the use of Machine-Learning and Artificial Intelligence by educational portals and OER repositories (white paper), 2018.

[CL07] Cuthrell, K.; Lyon, A.: Instructional strategies. What do online students prefer. In Journal of Online Learning and Teaching, 2007, 3; pp. 357–362.

[CMS17] Chatti, M. A.; Muslim, A.; Schroeder, U.: Toward an open learning analytics ecosystem: Big data and learning analytics in higher education. Springer, 2017; pp. 195–219.

[CP12] Chaffey, D.; Patron, M.: From web analytics to digital marketing optimization. Increasing the commercial value of digital analytics. In Journal of Direct, Data and Digital Marketing Practice, 2012, 14; pp. 30–45.

[CV13] Chrysafiadi, K.; Virvou, M.: Student modeling approaches. A literature review for the last decade. In Expert Systems with Applications, 2013, 40; pp. 4715–4729.

[Dr15] Drachsler, H. et al.: Panorama of recommender systems to support learning: Recommender systems handbook. Springer, 2015; pp. 421–451.

[Ga07] Gauch, S. et al.: User profiles for personalized information access: The adaptive web. Springer, 2007; pp. 54–89.

[Jo09] Jones, A.: Redisciplining generic attributes. The disciplinary context in focus. In Studies in Higher Education, 2009, 34; pp. 85–100.

[Ke11] Kent, M. L. et al.: Learning web analytics. A tool for strategic communication. In Public Relations Review, 2011, 37; pp. 536–543.

[MDK11] Magnisalis, I.; Demetriadis, S.; Karakostas, A.: Adaptive and intelligent systems for collaborative learning support. A review of the field. In IEEE Transactions on Learning Technologies, 2011, 4; pp. 5–20.

[ND08] Nguyen, L.; Do, P.: Learner model in adaptive learning. In World Academy of Science, Engineering and Technology, 2008, 45; pp. 395–400.

[RL08] Razmerita, L.; Lytras, M.D. Eds.: Ontology-based user modelling personalization. Analyzing the requirements of a semantic learning portal. Springer, 2008.

[Sp00] Specht, M. Ed.: ACE-adaptive courseware environment. Springer, 2000.

[Ve12] Verbert, K. et al.: Context-aware recommender systems for learning. A survey and future challenges. In IEEE Transactions on Learning Technologies, 2012, 5; pp. 318–335.

[Wi07] Willems, J. Ed.: Does style matter? Considering the impact of learning styles in e-learning. Australasian Society for Computers in Learning in Tertiary Education, 2007.

[ZP15] Zheng, G.; Peltsverger, S.: Web analytics overview: Encyclopedia of Information Science and Technology, Third Edition. IGI Global, 2015; pp. 7674–7683.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 187

# A mobile campus application as a sensor node for Personal Learning Environments

**Hybrid cross-platform User Contextualization**

Hendrik Geßner[1] (0000-0002-7786-2587) and Alexander Kiy[1] (0000-0002-5885-9657)

**Abstract:** For many web-based applications, there is at least one corresponding mobile application. By leveraging a mutual exchange of the cross-device user context between mobile and web-based application, guidance and processes can be improved and therefore simplified. This article presents an infrastructure to enable cross-device and cross-service personalization and adaption while aiming at high interoperability between heterogeneous systems. As a proof-of-concept, an existing mobile campus app framework was extended by a hybrid context framework to capture user data, which is stored in a Learning Record Store (LRS) by the use of the Experience API.

**Keywords:** mobile application, personal learning environment, xAPI, informal learning, context sensitivity

## 1 Motivation

While a few years ago there were only a few device classes that played a role in student life, there has been a rapid increase in device diversity. With full wireless coverage at universities coming into reach, the expansion of further mobile services to support teaching and learning is still at the beginning. To meet the individual challenges, a variety of services and mobile applications are used. As part of the university, formal and informal teaching and learning contexts are becoming increasingly blurred so that there can be no comprehensive use of information and data. The different contextual user data offers the opportunity to adapt the acquired institutional teaching and learning environment to the individual needs of the users in the best possible way. In the area of HCI, a large number of model- and rule-based variants for adapting mobile user interfaces already exist [FL15]. So-called context frameworks are increasingly used, which make it possible to configure both the mobile sensors used and the rules applied. However, these solutions are usually limited to the mobile device. On the other hand, the cross-device and multi-service capture and use of context information requires an underlying infrastructure. Hu et al. proposed a context aware architecture for mobile crowd sensing using a mobile SOA framework for mobile devices and a dedicated server component [HLN+14]. The approach makes it possible to capture context data across different mobile devices and to adapt the application to the device. However, the data management is decentralized so that

[1] University of Potsdam, Department of Computer Science, August-Bebel-Str. 89, 14482 Potsdam, {givenname.surname}@uni-potsdam.de

doi:10.18420/delfi2019_356

contextual data cannot be reused for other applications. This article introduces an infrastructure and a proof of concept that enables the use mobile context information for the adaptation of a web-based application. The hybrid mobile application Mobile.UP is extended by a context framework and the mobile context information is used for adaptation in the existing web-based personal learning environment Campus.UP [KLL17]. Thus it is conceivable that the last work status can be seamlessly resumed after a device change (mobile application to corresponding web application) and sensor information can also be used to enrich other services and mobile applications.

## 2 Conceptualization of the sensor node for PLEs

In order to find suitable use cases for the prototype implementation, we examined work from three different adjacent areas, collected ideas as user stories and enriched these with required context data and portlets before selecting two for prototype implementation. Economides [Eco09] provides an extensive list and classification of context sources in the field of pervasive and ubiquitous learning and Scheffel et al. [STD16] present a xAPI vocabulary. Mobile.UP [KLL17] contains features suitable for context extension. These sources resulted in two user stories[2]. User story 1, *current course*, states that if the user is currently attending a course his or her PLE should highlight the courses workspace. User story 2, *profile completion*, states that when completing the personal profile, the user should be supported by his or her PLE through suggestions[3].

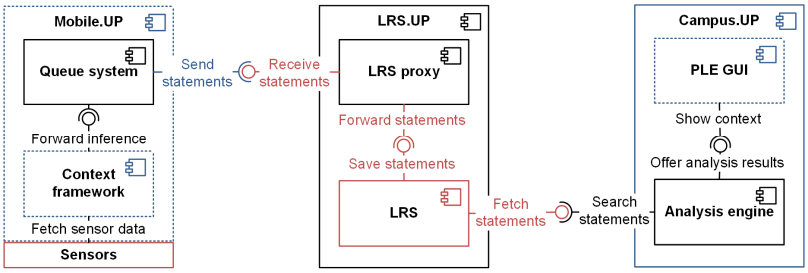



Fig. 1: System architecture with standardized components (red) and internal developments (blue).

The architecture consists of an extension of the mobile application Mobile.UP with a context framework for the acquisition of sensor information and the subsequent processing. Then the context information is transformed into xAPI statements. To handle changeable network connections, the application can be operated largely without a network, as a queue system sends the information to a central storage location, the so-called Learning Record Store (LRS), via the xAPI-specified interface [ADL16] as soon as possible. In

[2] A complete list of use cases can be accessed at https://dx.doi.org/10.5281/zenodo.3263934
[3] However, the evaluation of user story 2 showed that the implemented profile suggestions made no difference in efficiency, effectivity or user experience while only a small portion of suggestions were adopted at all. This led to the decision to omit any further details on the second use case in this article.

order to draw conclusions from the data stored in the LRS, an analysis component is used which contains knowledge about the structure and context of the data. As part of the previously selected use case *current course*, the display in the PLE is extended by additional information which is retrieved by the analysis engine via REST and then presented in the PLE. Taking into account the existing components, the architecture presented in figure 1 was designed (dashed lines symbolize the extension of an existing framework or system).

## 3 Implementation

The embedded *sensor and context framework* transforms the low-level context of the sensors into high-level context, which is then transformed into xAPI statements. The embedding of the framework allows access to smartphone and user-specific data and enables decentralized, network-independent processing. Several existing context frameworks were considered [Bar05, FKD15, MFN+13, MLL15, SDA99, Win17]. To ease integration into the Mobile.UP app, a mobile hybrid framework with offline capability and Cordova compatibility was required. This only left contactJS [MLL15], a JavaScript port of the Context Toolkit framework.

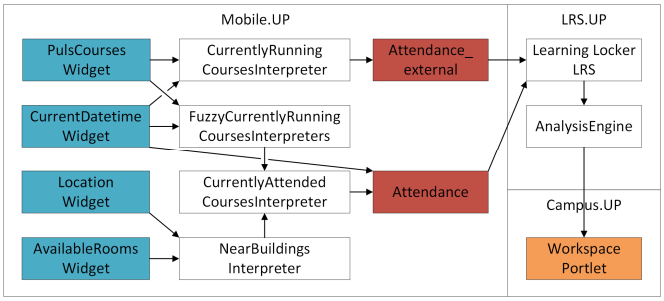



Fig. 2: Component-wise information flow from context framework to workspace highlighting

ContactJS employs so-called widgets to capture the low-level context, interpreters for further processing potentially low-level context to high-level context and rules to trigger conditional actions. In total, eight widgets, five interpreters and four rule sets were implemented of which a selection is presented in figure 2. The widgets retrieve building data as GeoJSON objects, current time in ISO 8601, device position following the HTML5 Geolocation API and enrolled user courses in a custom format, which is returned by the campus management system's web service. By matching high-level context with stored rules, xAPI statements are generated and sent to the *LRS*. The Experience API (xAPI) [ADL16] is a standard for capturing learning activities. For reusable scenarios, recipes exist which define the permitted vocabulary [BSD+16]. The communication between

mobile app and LRS follows the *ATTENDANCE* recipe[4][5]. To select an appropriate LRS implementation, the xAPI adopter list[6] was used. Mandatory criteria were free availability and the possibility of a local stand-alone installation. Only Learning Locker and IxHive fulfilled the requirements with Learning Locker providing an advanced query interface. As a result, Learning Locker was chosen. Communication between *analysis engine* and LRS was initially done through the xAPI query interface and the provided xAPI Java library. However, in addition to the intended statements, the LRS response also contained invalidated statements, their invalidation, and referenced statements. This behavior is correct with regards to xAPI, but unfavorable. Therefore, it was decided to use Learning Locker's aggregation API, which proved to be much more flexible and reliable. The analysis engine supports a requests for current user courses. Currently attended courses are determined by querying the LRS for all xAPI statements that match the detailed attendance recipe grouped by event ID and replaying the actions. The existing *PLE* Campus.UP is based on the open source portal software Liferay. The adjustment of features can be done via Hooks or by creating self-contained Portlets. Two existing Portlets, the workspace overview and the portlet for user profiles, were extended as part of this project. Highlighting the current courses workspace is performed by leveraging the already existing workspace overview portlet and extending it to regularly poll the *current course* REST interface and highlight designated workspaces in red.

## 4 Evaluation

The evaluation of the *current course* use case maintains a technical view by focusing on the time it took until the workspace linked to the attended course was highlighted in the PLE. An analytical evaluation was done, which yielded best and worst case intervals. Afterwards, these bounds were used to interpret the recorded durations collected from several experimental runs. Two starting points were covered: The user already arrived at the venue 10 minutes before the event start (starting point 1) OR the user is on his or her way to the venue at the 10-minute-mark (starting point 2). The starting point affects the expected system reaction times and delays. Figure 2 highlights the time-critical components. *CurrentDatetimeWidget* refreshes every 30 seconds, *Attendance* rules are rechecked in a 20 second interval and the Workspace portlet queries for new data every 30s. The *LocationWidget* refreshes every 60s but depends on the underlying locations plugin, whose refresh rate is not predictable due to mechanisms like geo fencing. The data processing and transfer happens in a millisecond range and will therefore be neglected in the analysis. Maximum runtime is reached when each component has to wait a full refresh cycle before processing prior results. In starting point 1, this results in a remaining theoretical worst-case runtime of 30s + 20s + 30s = 80s. The experimental runs yielded runtimes of 22s, 40s and 50s, the average case is 37.3s. All values are below the analytical

[4] https://registry.tincanapi.com/#profile/48
[5] A list of implemented xAPI statements can be accessed at https://dx.doi.org/10.5281/zenodo.3263934
[6] https://xapi.com/adopters/

maximum or worst case of 80s which strengthens the value of 80s as the worst case limit. In practice, these values denote that in an implementation which may recognize a course as active 10 minutes before it starts, the associated workspace is highlighted no later than 8.5 minutes before it starts. A maximum delay of 1.5 minutes can be seen as acceptable or good. The theoretical worst-case maximum runtime in starting point 2 includes the *LocationWidget* refresh interval. This results in max(30s, 60s) + 20s + 30s = 110s, but may be higher due to the underlying location plugin. In the experimental evaluation, the test device was taken to a location about 2.3 km away from the designated event venue well before the 10 minute mark. Upon re-arrival, the time until highlighting of the expected course was measured and the timestamp of the *joined* statement sent by the test device was noted. Two runs were done which yielded runtimes of 0s each and *joined* statement send timestamps of 62s and 102s before venue arrival. The average case is 0s with a statement sent 82s before arrival. In practice, this means that an implementation which updates its device position every 60s can highlight the associated workspace immediately on arrival at an event venue. This represents the optimal result for starting point 2.

## 5 Conclusion & Outlook

An architecture for cross-system data gathering, processing, transfer and usage of context information by using standardized and conventional technologies was presented. The user information gathered by the app were utilized to automatically adapt the PLE. For that purpose two representative use cases were selected and implemented prototypically. The designed architecture bases on standards like xAPI while remaining flexible enough to allow additional use cases. The prototypes evaluation showed that it is very suitable to highlight current courses in the PLE. The worst case and average case time limits result from refresh intervals within the app and the PLE. One can expect to reduce these limits by implementing higher refresh rates. Nevertheless, one should keep in mind that higher refresh rates come at the expense of increased power consumption and data traffic, which are precious resources on mobile devices systems such as smartphones and laptops. Cordova cannot guaranty JavaScript background execution, which is necessary for continuous context gathering and processing. Additionally, there is no prior experience on the power consumption of a hybrid Cordova app when compared to a native solution. In each case a native approach may be more suitable. As part of this prototype it has been shown that a number of auxiliary functions are required (cf. xAPI *opened* statement) that could be provided by other infrastructure (e.g. campus or learning management system) via active push of user context information into the LRS. The increased crosslinking of systems is now so far that fields like Learning Analytics rearrange themselves because services linked by standards like xAPI yield more detailed and more comprehensive insights than ever before. There is hardly any long-term experience and no broad consensus regarding cross-platform and cross-service vocabulary [BSD+16].[7]

[7] The research of Hendrik Geßner has been partially funded by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1294 “Data Assimilation”, Project Z03 “Information Infrastructure for data assimilation”.

## Bibliography

[ADL16] Advanced Distributed Learning (ADL) Initiative, U.S. Department of Defense (2016) xAPI Specification. https://github.com/adlnet/xAPISpec

[Bar05] J. E. Bardram: The java context awareness framework (JCAF) - a service infrastructure and programming framework for context-aware applications. PERVASIVE'05 Proc. of the 3rd Int. Conf. on Pervasive Computing. Springer-Verlag, Berlin, Heidelberg, pp. 98-115, 2005. doi:10.1007/11428572_7.

[BSD+16] A. Berg; M. Scheffel; H. Drachsler; S. Ternier & M. Specht: The dutch xAPI experience. In: (D. Gašević, G. Lynch, S. Dawson et al., Eds.) LAK '16 conference proceedings: The 6th Int. Learning Analytics & Knowledge Conference, The University of Edinburgh. The Association for Computing Machinery, New York, pp. 544–545, 2016.

[Eco09] A. A. Economides: Adaptive context-aware pervasive and ubiquitous learning. IJTEL 1(3): pp. 169-192, 2009. doi:10.1504/IJTEL.2009.024865.

[FL15] J. Feng and Y. Liu: Intelligent context-aware and adaptive interface for mobile LBS. Journal. 2015.

[FKD15] D. Ferreira; V. Kostakos & A. K. Dey: AWARE: Mobile Context Instrumentation Framework. Frontiers in ICT. 2, 2015. doi: 10.3389/fict.2015.00006

[HLN+14] X. Hu; X. Li, E. Ngai; V. Leung & P. Kruchten: Multidimensional context-aware social network architecture for mobile crowdsensing. IEEE Communications Magazine. 46(6): pp. 78-87, 2014. doi:10.1109/MCOM.2014.6829948.

[KLL17] A. Kiy; C. List and U. Lucke: A virtual environment and Infrastructure to ensure future readiness of Data Centers. European Journal of Higher Education IT (EJHEIT), 2017.

[MFN+13] M. E. F. Maia; A. Fonteles; B. Neto; R. Gadelha; W. Viana & R. M. C. Andrade: LOCCAM - loosely coupled context acquisition middleware. In: Proc. SAC '13 Proc. of the 28th Annual ACM Symposium on Applied Computing. ACM New York, NY, USA, pp. 534-541, 2013. doi:10.1145/2480362.2480465.

[MLL15] T. Moebert; S. Lemcke & U. Lucke. ContactJS -- A Cross-Platform Context Detection Framework. In: IEEE 15th Int. Conf. on Advanced Learning Technologies, pp. 108-110, 2015. doi:10.1109/ICALT.2015.83.

[SDA99] D. Salber; A. K. Dey & G. D. Abowd. The context toolkit: aiding the development of context-enabled applications. CHI '99 Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. pp. 434-441, 1999. doi:10.1145/302979.303126.

[STD16] M. Scheffel; S. Ternier & H. Drachsler: The Dutch xAPI Specification for Learning Activities (DSLA) – Registry, 2016. http://bit.ly/DutchXAPIreg

[Win17] J. Winterfeldt: Context-Aware Mobile Crowd Sensing using Mobile Hybrid Application Frameworks. Masters thesis. Universität Ulm, 2017. http://dbis.eprints.uni-ulm.de/1486/

# Sensor Based Adaptive Learning - Lessons Learned

Albrecht Fortenbacher[1], Manuel Ninaus[2], Haeseon Yun[3], René Helbig[4] and Korbinian Moeller[5]

**Abstract:** Recent advances in sensor technology allow for investigating emotional and cognitive states of learners. However, making use of sensor data is a complex endeavor, even more so when considering physiological data to support learning. In the BMBF-funded project *Learning Analytics for sensor-based adaptive learning (LISA)*, we developed a comprehensive solution for adaptive learning using sensor data for acquiring skin conductance, heart rate, as well as environmental factors (e.g. $CO_2$). In particular, we developed, (i) a sensor wristband acquiring physiological and environmental data, (ii) a tablet application (SmartMonitor) for monitoring and visualizing sensor data, (iii) a learning analytics backend, which processes and stores sensor data obtained from SmartMonitor, and (iv) learning applications utilizing these features. In an ongoing study, we applied our solution to a serious game to adaptively control its difficulty. Post-hoc interviews indicated that learners became aware of the adaptation and rated the adaptive version better and more exciting. Although potentials of utilizing physiological data for learning analytics are very promising, more interdisciplinary research is necessary to exploit these for real-world educational settings.

**Keywords:** sensor based learning, learning analytics, adaptive learning system

## 1 Introduction

The current wave of digitization has enormous potential to impact education and training. Consider, for instance, the situation of teachers. The nature of teaching, especially in schools and universities, has not changed significantly for more than 200 years. However, the prevalent type of direct instruction rarely allows individual learning states to be addressed. Digital learning environments can help collecting data at the individual learner level and providing (semi-)automated feedback on indicators for cognitive and emotional states such as mood or concentration [Sb15]. This may help to identify students' individual needs, but it also may allow for giving targeted feedback. Learning analytics – utilizing data from digital learning environments appropriately to understand and support learning processes – becomes particularly challenging when going beyond conventional user interaction data by considering data provided from physiological sensors such as heart rate, skin conductance response, and others. Sensor data are particularly interesting as they allow for deeper insights into individual cognitive and emotional states crucial for

[1] A. Fortenbacher, HTW Berlin, Wilhelminenhofstraße 75A, 12459 Berlin, forte@htw-berlin.de
[2] M. Ninaus, Leibniz Inst. für Wissensmedien, Schleichstr. 6, 72076 Tübingen, m.ninaus @iwm-tuebingen.de
[3] H. Yun, HTW Berlin, Wilhelminenhofstraße 75A, 12459 Berlin, yun@htw-berlin.de
[4] R. Helbig, HTW Berlin, Wilhelminenhofstraße 75A, 12459 Berlin, helbig@htw-berlin.de
[5] K. Moeller, Leibniz Inst. für Wissensmedien, Schleichstr. 6, 72076 Tübingen, k.moeller@iwm-tuebingen.de

doi:10.18420/delfi2019_355

learning. Consequently, increasing research interest has been devoted to the role of learning analytics considering physiological data [Sb15]. In the BMBF-funded project *Learning Analytics for sensor-based adaptive learning (LISA)*[6], we pursued sensor based adaptive learning in an interdisciplinary approach. When starting to think about sensor-based support for individual learning in LISA, some questions arose: Which sensors should be used for a successful learning solution? However, which of them are suitable to support self-regulated learning [YFP17]? And which sensors are easily available and suited for non-intrusive and non-distracting learning support? Moreover, even when the most appropriate (physiological) sensors were chosen, further questions arise with respect to how to process the acquired data by means of sensor based learning analytics. Can cognitive and emotional states of learners be measured and predicted reliably? How can a sensor based system become a learning companion? Which user centric design should be used? Is such a system capable of giving empathetic feedback and/or recommendations? In the following sections, we will refer to some of these questions in the context of our research project.

## 2 From Sensor Data to Learning Indicators

Among different sensor data, measures derived from electrodermal activity (EDA), electrocardiogram (ECG) or photoplethysmogram (PPG) were often found to be sensitive to emotional states [Kre10]. Furthermore, combinations of measures derived from EDA and ECG were used to describe different dimensions of emotions [Gr15], [MO13]. To derive learning indicators from these sensor data, practical approaches such as machine learning were previously used to classify emotions, and it was shown that aggregated sensor data could reflect emotions reliably [PVH01].

Requirements for the selection of a LISA sensor device were, amongst others: 1) utilization of sensors indicative of emotional and cognitive states, 2) easy application in real learning situations, 3) data privacy. From these requirements, EDA and PPG sensors were chosen, as well as sensors for skin temperature, total volatile organic compounds (TVOC) and $CO_2$. Using the developed LISA sensor device, we successfully evaluated emotion detection using different methods (e.g. qualitative, quantitative, machine learning and fuzzy logic approaches, see [Yu17, Yu19]). Additionally, we are on the way to also be able to indicate cognitive states of learners, in addition to their emotional states.

## 3 Bringing all parts together

A LISA learning application consists of 4 components: 1) sensor wristband, which records EDA, PPG, temperature, TVOC, $CO_2$ data, 2) tablet application (SmartMonitor), which processes sensor data and acts as a learning companion, visualizing sensor data and

[6] Förderkennzeichen 16SV7534K, see http://tel.f4.htw-berlin.de/lisa

providing awareness/feedback to the learner, 3) LISA learning analytics backend, which receives sensor data from SmartMonitor, stores learner data and provides learning analytics services to 4) learning applications (see Fig. 1).

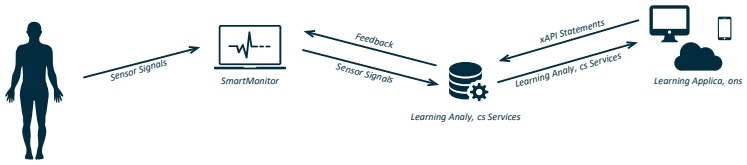



Fig. 1: Illustration of sensor-based learning as a distributed application

To provide multimodal learning analytics services, the LISA data model can hold learner activities and streams of raw sensor data, but also aggregated sensor data like heart rate (variability) or timestamps when e.g. heart rate exceeds critical values. Multimodal learning analytics services range from predicting emotional and cognitive states to detecting learning situations. In addition to providing a complete learning analytics solution to learners, data privacy and ethical issues were investigated to ensure learner's ownership and agency over their data. In LISA, sensor data can only be used under the strict permission of the learner. For this purpose, a model of trust has been developed in LISA which controls the usage of data by user preferences. To achieve this goal, a learning application provides a token which a learner can use to establish a learning session, connecting the personal SmartMonitor device to the respective learning application. Thus, sensor data can be mapped to learner IDs without the need to store any information about a learner in SmartMonitor or LISA backend

## 4 Sensor-based Learning Applications

### 4.1 LISA Learning Companion

A first adaptive learning solution developed under the LISA project is a learning companion which promotes self-regulated learning. The learning companion is a human-like interactive peer which utilizes sensor data to provide learners with a positive learning experience along with self-regulated learning support [YFP17]. The learning companion gains information about learners' mental states by using physiological and environmental data. Based on analyses of these sensor data, both alerts (e.g., exceeding a critical threshold in air quality) and interpreted recommendations (e.g., volitional control strategies) were designed. In LISA, we considered human computer interaction design and technical feasibility along with pedagogical concepts (Fig. 2). The LISA learning companion enriches learners' learning experience by providing an opportunity to set and also track their learning goals (an active learning goal can be seen on the left side of Fig. 2). While learning, learners can also attain volitional control strategies like self-efficacy enhancement or stress reducing actions [MG99]. To make interaction with the learning

companion more enjoyable, learning related humors and motivational quotes have been chosen from online resources. The first field test with a group of students who used the learning companion provided promising feedback. In particular, the learning companion was considered *sweet* and *cool* and reported to keep learners from distraction. The results also showed that it engaged learners to reflect on their learning and allowed them to apply the current reflected experience to plan their next learning more effectively.

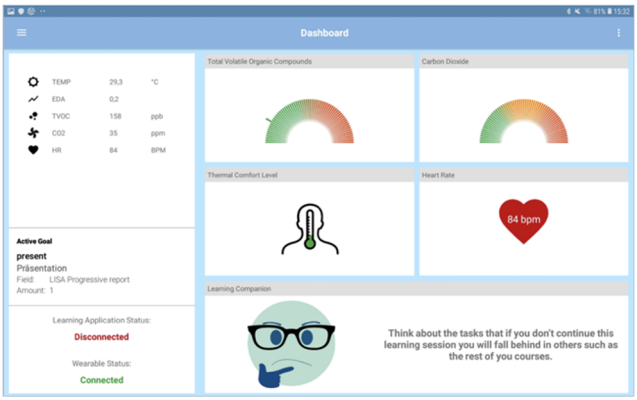



Fig. 2: Screenshot of LISA Learning Companion

### 4.2 An Adaptive Learning Game

In the project, we developed an adaptive version of *Emergency*, a serious game by Promotion Software[7], as one potential application scenario for sensor based learning. In this game players coordinate firefighters and rescue services in different scenarios such as train and car crashes to save and patch up wounded characters and prevent houses and vehicles from burning down. The difficulty of the game is related to the number of tasks that need to be addressed simultaneously, which creates stressful situations. For instance, with increasing difficulty, more wounded characters had to be cared for concurrently, or the risk of fire was higher. Hence, adaptation of the level of difficulty within the game seems desirable for keeping learners in the "sweet spot" of the Yerkes–Dodson Law [YD08] postulating an inverted U-shaped pattern for the interrelation between performance and arousal: individuals' maximum performance should be achieved on a medium level of arousal whereas performance should decrease when the arousal becomes too low or too high.

For adapting the game's difficulty level, we used the LISA SmartMonitor with a particular focus on heart rate data due to its close link to stress and arousal [Je11]. Moreover, previous pilot studies indicated that heart rate is linked to the difficulty levels of the game. Therefore, we developed a mechanism adapting the difficulty based on the player's

[7] https://promotion-software.de/

individual heart rate. During an initial heart rate perception phase, we measured the resting pulse of a player to automatically implement a baseline and threshold values (i.e., 5 beats above or below resting pulse) for triggering adaptations in the game. For instance, given a baseline of 75 beats per minute (BPM), the game became easier when heart rate exceeded 80 BPM, whereas it became more difficult when BPM fell below 70 BPM. To support overstrained players, helicopters appeared and automatically patched up patients so that the player no longer had to care for them. In contrast, when arousal was low, bystanders appeared at accident sites and got hurt, which increases the number of patients to be cared for by the player.

In an ongoing piloting study, we examine the feasibility of adapting the game's difficulty based on changes in players' heart rate. So far, 12 participants ($M_{age}$=28.83 years; *SD*=5.25 years) not aware of the experimental manipulation played both an adaptive and a non-adaptive version. Initial results of interviews after the playing sessions focusing on user experience revealed that participants became aware that the game became easier/more difficult in the adaptive version. Interestingly, participants also rated the adaptive version to be better (58%), more exciting (75%), but also more demanding (83%), indicating feasibility of the current implementation of an adaptation based on physiological data.

## 5 Lessons Learned and Future Considerations

Our endeavor to consider sensor data in a context of learning started ambitiously, and we included psychological, technical and real-life learning considerations into our research. Using sensor data for learning support requires decisions on a lot of questions, e.g. statistical significance vs. ecological validity, data ownership vs. learners' agency, human computer interaction design vs. pedagogical support. From our research we learned that quick and easy results are often neither realistic nor meaningful. In particular, the process of considering sensor data for learning analytics is by no means straightforward. It entails complex analysis and modification of signals to provide information about behavior and/or states of a user. Using user interaction data for learning analytics is also complex and becomes even more challenging when physiological data are used in learning scenarios to detect emotional, motivational, and cognitive states. The reason is that physiological data provide indicators for such mental states which cannot be measured directly. For instance, changes in heart rate as observed in players of the *Emergency* game may indicate emotional arousal but may also be an indicator for cognitive effort as well as motoric activity, or in the worst case just noise from the sensor. Nevertheless, the potential for user centered adaptation of learning environments is immense, and collaborative interdisciplinary research involving various stakeholders (e.g. universities, research institutes, industry partners) – as pursued in LISA – will allow mastering such complex research and applying real-life questions successfully.

## Bibliography

[AC04] Azevedo, Roger; Cromley, Jennifer G: Does training on self-regulated learning facilitate students' learning with hypermedia? Journal of educational psychology, 96(3):523, 2004.

[Gr15] Gruber, J.: Mennin, D. S.; Fields, A.,; Purcell, A.; Murray,G: Heart rate variability as a potential indicator of positive valence system disturbance: a proof of concept investigation. International Journal of Psychophysiology, 98(2):240–248. 2015.

[Kre10] Kreibig, S. D: Autonomic nervous system activity in emotion: A review. Biological psychology, 84(3):394–421. 2010.

[MG99] McCann, Erin J; Garcia, Teresa: Maintaining motivation and regulating emotion: Measuring individual differences in academic volitional strategies. Learning and individual differences, 11(3):259–279, 1999.

[MO13] Malkawi, Mohammad; Omayya Murad: Artificial neuro fuzzy logic system for detecting human emotions." Human-Centric Computing and Information Sciences 3, no. 1: 3, 2013.

[PVH01] Picard, R.W., Vyzas, E., and Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE transactions on pattern analysis and machine intelligence, 23(10):1175–1191. 2001

[Sb15] Schneider, J., Börner, D., Van Rosmalen, P., & Specht, M.: Augmenting the senses: A review on sensor-based learning support. Sensors (Switzerland), 15(2), 4097–4133, 2015.

[YD08] Yerkes, R. M., & Dodson, J. D.: The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 18, 459–482, 1908.

[Je11] Jerritta, S., Murugappan, M., Nagarajan, R., & Wan, K.: Physiological signals based human emotion recognition: a review. In 2011 IEEE 7th International Colloquium on Signal Processing and its Applications (pp. 410-415). IEEE, 2011

[YFP17] Yun, Haeseon; Fortenbacher, Albrecht; Pinkwart, Niels: Improving a Mobile Learning Companion for Self-regulated Learning using Sensors. In: CSEDU (1). S. 531–536, 2017.

[Yu17] Yun, Haeseon; Fortenbacher, Albrecht; Pinkwart, Niels; Bisson, Tom; and Moukayed, Fadi: A pilot study of emotion detection using sensors in a learning context: Towards an affective learning companion. In Ullrich, C. and Wessner, M., editors, Proceedings der Pre-Conference-Workshops der 15. E-Learning Fachtagung Informatik co-located with 16th e-Learning Conference of the German Computer Society (DeLFI 2017), volume 2092. CEUR Workshop Proceedings (CEUR-WS.org ). 2017.

[Yu19] Yun, Haeseon; Fortenbacher, Albrecht; Helbig, René; Pinkwart, Niels: In Search of Learning Indicators: A Study on Sensor Data and IAPS Emotional Pictures. In: CSEDU (in press), 2019.

# Code of Practice for Sensor-Based Learning

Haeseon Yun[1], Shirin Riazy[2], Albrecht Fortenbacher[3] and Katharina Simbeck[4]

**Abstract:** Sensor-based learning refers to utilizing physiological sensor data from learners and information from a learning environment to promote learning. Sensor data enclose learner's personal information so ethical practice of adopting sensor data in learning analytics needs to be explored thoroughly. In this positional paper, we examine current ethical practices in learning analytics to derive a code of practice for sensor-based learning. Furthermore, we critically validate a wearable sensor device developed as a learning support against the derived code of practice.

**Keywords:** Learning Analytics, Sensor based learning, ELSI, GDPR, Sensor Data

## 1 Introduction and Motivation

The introduction of "General Data Protection Regulation (GDPR)" leads to an overall raised awareness of sensible data being collected and processed. Not only organizations like enterprises or universities are affected, but also individual users are prompted to consent for their data being transferred and utilized.

In order to understand learning conditions for instructors and students and to provide learners with the most appropriate context-aware feedback, learning analytics uses digital footprints in a learning environment. Even though the intention of learning analytics is noble, it is important to investigate the available measures and framework which ensure the proper usage of data. In the case of sensor-based learning support, where physiological data is used, GDPR provides an overall guideline of data privacy to protect the freedom and the rights of a person. Specifically, it should be prohibited to use any genetic or physiological data which can identify a person or can be tracked back to a user. Furthermore, any analysis and prediction of "personal aspects" to create a "personal profile" is banned [EP16]. In addition to general ethical guidelines such as the Nuremberg Code [Sh97] or the Declaration of Helsinki [As01], various researchers introduced principles [SP13], frameworks [Wi14] and checklists [DG16] in the context of learning analytics. Recently, there have also been concerns on how to deal with data-driven interventions, generated by machine learning and big data algorithms [Wi14].

Under the three-year project LISA [5] we have researched the feasibility and the implementation of sensors in a learning support system using machine learning

[1] University of Applied Science, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, yun@htw-berlin.de
[2] University of Applied Science, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, riazy@htw-berlin.de
[3] University of Applied Science, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, forte@htw-berlin.de
[4] University of Applied Science, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, simbeck@htw-berlin.de
[5] http://tel.f4.htw-berlin.de/lisa, funded by German Ministry of Education and Research, 16SV7534K

doi:10.18420/delfi2019_326

algorithms. Among various research aims, ELSI (ethical, legal, and social implications) was one major objective of the LISA project when collecting and utilizing students' data including sensor data.

In complex systems, where learner's biographical, behavioral and physiological data are used, serious and adamant reviews and critiques are necessary. Therefore, in this positional paper, we have 1) reviewed previous related research in data privacy, legal, ethical and social issues in learning analytics, 2) derived and discussed the core code of practice for sensor-based learning support and 3) critically reviewed the case of a sensor-based learning support system using the derived code of practice.

## 2 Derived Core Code of Practice

Based on our literature review of ethical guidelines, frameworks, principles and future considerations in the context of learning analytics, we derived six codes of practice, which are summarized in Tab. 1.

| **Code of Practice** | **Brief Explanation** | **Reference** |
|---|---|---|
| Consent & Transparency | Consent of subjects (learners and teachers) and transparency of measurements and processes are required. | [SP13, Sh97, As01, Fe16, SB15, SB14, DG16, WM15, PS17, EP16] |
| Data Ownership | Student should act as an active agent. Data ownership, security of data storage and management should be clear. | [Fe16, Wi14, PS17, SP13, SB14, SB15, Si13, Li16, DG16, As01] |
| Algorithm & Interventions | Interventions should aim at positive learning. Researchers should be conscious of the purpose and effects of interventions. Algorithmic results should not be perceived as a ground truth. | [Wi14, EP16, SB15, SB14, SP13, PS17, Wi16] |
| Privacy | Privacy of all subjects should be protected by involving all stakeholders continuously. | [Fe16, SB15, DG16, Si13] |
| Legal Responsibility | Researchers should be aware of legal conduct and responsibilities. Data used for learning analytics should be accurate and up-to-date. | [SP13, Fe16, SB15, SB14, As01, DG16, Sc16] |
| Higher Standards | Avoid any mental and physical harm. Humanitarian benefits should play a central role. Higher ethical standards should be applied. | [Fe16, Sh97, As01, SP13, SB15, SB14, DG16, PS17] |

Tab. 1: Overview of Code of Practice points for ethical research grouped by topic.

## 2.1 Consent & Transparency and Data Ownership

The first code of practice, consent and transparency, refers to making the usage of data by institutions clear. This includes specifying the objectives and purposes of the scope and context where data is used including collection, analysis and utilization [SB15, SB14, DG16, SP13, WM15]. When personal intervention is required, this practice implies requiring students' voluntary consent [As01, DG16, WM15]. However, as indicated by GDPR [EP16], institutes should not rely only on user's consent when processing users' data. Instead, they should consider other issues that might infringe user's rights.

Under the code of practice "data ownership", the handling of the data should engage students as "active agents" and as "collaborators" in the implementation, interventions and services of learning analytics [SP13, SB14]. In practice, this should be realized using Engelfriet's three steps ("easy access" to collected data, "the right to correct wrong information and interpretations" and "the right to remove irrelevant data") [SP13]. Additionally, it is crucial that students, as actors in learning analytics, are able to opt out from data processing [DG16, PS17, As01]. For sensor data, there have been attempts to enable transparent and integrative data handling, such as in the RECOLA Database[6] and MyData Button [Li16]. However, challenges still remain when attaining students' consent, as various reasons and factors could affect the decisions of students [PS17].

## 2.2 Algorithms & Interventions

To provide positive interventions, models and algorithms being used should be "sound and free from bias" ([SB14, SB15]). Specifically, Prinsloo and Slade emphasized the special care when applying artificial intelligence on educational data as "they can perpetuate, exacerbate, or mask harmful discrimination" [PS17]. To mitigate the negative effects of algorithm applied in educational data, the usage of algorithms should consider "how algorithms reinforce, maintain, even reshape visions of the social world, knowledge and encounters with information" as the application of learning analytics on educational data would "evaluate and manage corporeal, emotional and embrained lives" [Wi16, PS17]. For learning analytics interventions, one should consider that evaluations of sensor data could be biased by specific health conditions, race or gender. Student's identity and performance should be regarded as "temporal dynamic constructs" [SP13]. Various authors stated that learning analytics only provides a facet of students' behavior, therefore students' success should not be defined by the results of learning analytics but regarded as "a complex multidimensional phenomenon" [SP13, SB15, SB14]. Accordingly, it must be considered that sensor data reflects students' physiology during a comparably short time period of measurement and thus provides just a glimpse into emotional learning states.

[6] https://diuf.unifr.ch/main/diva/recola/index.html

### 2.3 Privacy and Legal Responsibility

[SB15] emphasized privacy as a distinctive area to consider when using learning analytics. Specifically, [DG16] proposed to 1) anonymize individual data so that they are not retrievable or trackable, 2) monitor and secure technical implementation to meet the privacy standard, 3) involve all stakeholders to secure the strict privacy measures and to be open to privacy concerns which might arise.

With respect to the legal responsibility, data processing should be fair and legal and reflect a legitimate purpose [As01]. Also, the people involved in learning analytics should be prepared for any legal consequences [DG16]. To assure the legal aspects, [Sc16] proposed the need for "extensive literature review of the legal and ethical issues around learning analytics". [SB15] emphasized the responsibility of institutions to take strict measures when dealing with students' data and providing interventions. Institutions should consult with student representatives and key members on issues of "objective, design, development, roll-out and monitoring of learning analytics" [SB15]. As sensors provide personal data, the data should both be accurate and up-to-date, and should adhere to a legal responsibility of institutions [EP16].

### 2.4 Higher Standards

The last derived code of practice refers to ethical principles which include both robust and sovereign purpose (greater good). Various authors state that this includes an inner attitude, namely opposing any harm to the people involved, and a regard for the society [SB15, DG16]. Based on the Nuremberg Code, the process of learning analytics should "avoid unnecessary physical and mental suffering" and it should consider any reasons of negative implications [Sh97]. Similarly, [As01] state that learning analytics should not put participating learners into any situations which cause either physical or psychological harm. To summarize, the design of a sensor-based learning support should specifically respect students' values and be in their interest [SB15, DG16, FM97].

## 3 Case of a Learning Support System Using Sensor Data

The LISA wearable sensor device includes both physiological sensors and sensors recording environmental data to provide self-awareness of learning states. Learners wear a LISA sensor device and interact with a learning support system which acts as a learning companion. It provides information about a learner's physiological state and about the learning environment, along with some basic and more sophisticated advice on demand.

With respect to consent & transparency and data ownership, using the sensor device may indicate automatic data collection and analysis, thus providing data ownership to the

service provider. For example, the sensor device Empatica[7], which is similar to the wristband developed in LISA, is known to collect biographical data such as name, date of birth and contact details along with all physiological and technical data. On the other hand, the LISA sensor device does not ask nor require users to provide any personal data and does not assume the data ownership. Learners' data is not transferred either to associated companies or to third parties who provide additional services nationally or internationally. Regarding algorithms & interventions, the LISA learning support system aims at providing learners with self-awareness by providing cognitive and emotional support so that learners can persist in learning. Here, privacy and transparency protect the users from potentially false interventions.

The LISA system does not record any personal data, and learners cannot be tracked by connected services using sensor data, to ensure learners' privacy. Even if learners' legitimate interests are shared by the service provider, our system requires learners' agency. Additionally, the development of LISA learning support system has been and currently is applying the higher standards by involving all stakeholders, including developers, engineers, researchers, instructors in higher education, HCI experts, communication designers, but especially students who should never be overlooked.

## 4 Discussion and Outlook

In this positional paper, we have reviewed previous studies that propose ethical guidelines in learning analytics. By consolidating research findings, we have derived a code of practice and conceptually applied it in the context of sensor-based learning. Then, we critically reviewed how the LISA learning support system fulfills our derived code of practice. The LISA system, in its current state, realizes a few codes of practice, namely consent & transparency, data ownership and privacy as opposed to commercially available sensor-based learning services. Our initiative to address concrete ethical issues of sensor-based learning analytics is in an early stage. However, as the critical review of sensor-based learning analytics is rare, our approach to derive a code of practice for sensor-based learning analytics is novel and necessary. Our next step will include thorough comparison of other sensor based systems to improve our code of practice.

Bibliography

[As01] Association, World Medical et al.: World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. Bulletin of the World Health Organization, 79(4):373, 2001.

[DG16] Drachsler, Hendrik; Greller, Wolfgang: Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In: Proceedings of the sixth international conference on learning analytics & knowledge. ACM, S. 89–98, 2016.

[7] https://www.empatica.com/privacy/

[EP16] European Parliament, Council of the European Union: , DIRECTIVE (EU) 2016/680 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016., 2016. Online; accessed 18 March 2019.

[Fe16] Ferguson, Rebecca; Hoel, Tore; Scheffel, Maren; Drachsler, Hendrik: Guest editorial: Ethics and privacy in learning analytics. SoLAR, 2016.

[FM97] Friedman, Batya; Millet, Lynette I: Reasoning About Computers as Moral Agents: A Research Note. In (Friedman, Batya, Hrsg.): Human values and the design of computer technology, Kapitel 11, S. 201–205. Cambridge University Press, 1997.

[Li16] Ligon, Glynn D.:MyData Button: Button, Button, Who's Got the Button? Nobody! Click one button and get all your education data! Too good to be true? Yes., 2016. Online; accessed 16 March 2019.

[PS17] Prinsloo, Paul; Slade, Sharon: Ethics and learning analytics: Charting the (un) charted. SOLAR, 2017.

[SB14] Slade, S; Boroowa, A: Policy on ethical use of student data for learning analytics. Milton Keynes: Open University UK, 2014.

[SB15] Sclater, Niall; Bailey, Paul: Code of practice for learning analytics. 2015.

[Sc16] Sclater, Niall: Developing a Code of Practice for Learning Analytics. Journal of Learning Analytics, 3(1):16–42, 2016.

[Sh97] Shuster, Evelyne: Fifty years later: the significance of the Nuremberg Code. New England Journal of Medicine, 337(20):1436–1440, 1997.

[Si13] Siemens, George: Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10):1380–1400, 2013.

[SP13] Slade, Sharon; Prinsloo, Paul: Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10):1510–1529, 2013.

[Wi14] Willis, JE: Learning analytics and ethics: A framework beyond utilitarianism. Educause Review, 2014.

[Wi16] Williamson, Ben: Coding the biodigital child: the biopolitics and pedagogic strategies of educational data science. Pedagogy, Culture & Society, 24(3):401–416, 2016.

[WM15] Welsh, Simon; McKinney, Stewart: Clearing the fog: A learning analytics code of practice. 2015.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 205

# Simplifying license attribution for OER with *emacs-reveal*

Jens Lechtenbörger[1]

**Abstract:** Open Educational Resources (OER) come with different license terms that require different forms of attribution. Properly attributing OER with licensing information has previously been identified as one of the most time-consuming factors of OER projects. As shown in this paper, the semantic annotation standard CC REL lacks features that are necessary for correct attribution with popular Creative Commons licenses. The software *emacs-reveal* addresses this gap in the context of OER slideshows, which may contain figures under different licensing terms: For each figure, a simple text file stores metadata building upon the vocabulary of CC REL (with pragmatic extensions) and can be shared along with the figure. From that metadata, the software generates attribution statements, both in machine-readable form (based on RDFa) and in human-readable form. The novel resulting process avoids manual copying of relevant license information, which is time-consuming and error-prone.

**Keywords:** Open Educational Resource, OER, Attribution, Creative Commons, RDFa, FLOSS

## 1 Introduction

According to the *Ljubljana OER Action Plan 2017* [Un17], "Open Educational Resources (OER) support quality education that is equitable, inclusive, open and participatory." A cornerstone of goals such as equity, inclusiveness, openness, and participation is proper licensing of educational resources. Indeed, following the *2012 Paris OER Declaration* [Un12], OER are "teaching, learning and research materials in any medium, digital or otherwise, that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions." Despite the widespread adoption of Creative Commons licenses for OER [AS16,ES18,Hi10], only few tools exist to help creators and users of OER with proper license attribution, which is among the most time-consuming factors for OER projects [Fa16].

This paper addresses challenges when creating, re-using, and mixing OER that are published under diverse licenses and presents the approach taken by the free/libre and open source software (FLOSS, see [Cr08] for a survey) *emacs-reveal* for proper license attribution of figures in slideshows. This software has been developed and used by the author since April 2017 to create OER presentations (HTML slideshows with audio explanations; to be viewed with standard Web browsers, either on- or offline). Users of *emacs-reveal* write source files for presentations in the lightweight markup language Org mode in the text editor GNU Emacs (see [SD11] for a general introduction to Org mode).

[1] ERCIS, WWU Münster, Leonardo-Campus 3, 48149 Münster, Germany, lechten@wi.uni-muenster.de

doi:10.18420/delfi2019_280

As sketched in [Le19], *emacs-reveal* is FLOSS that supports Single Sourcing [Ro01] of OER and meets the requirements of the ALMS framework by Hilton et al. [Hi10], which aims for the freedoms to reuse, redistribute, revise, remix educational resources; later Wiley [Wi14] added retain, leading to the "5 Rs of openness" of OER.

With *emacs-reveal*, figures can be accompanied by metadata files that specify licensing information based on standard Semantic Web vocabularies, notably the Creative Commons Rights Expression Language (CC REL) based on RDF [Ab12]. When using a figure in a presentation, proper human-readable license attribution is generated automatically from the corresponding metadata file and embedded in machine-readable RDFa format in the HTML presentation. An example for the generated HTML code is shown in Fig. 1, while Fig. 2 and Fig. 4 are screenshots of Browser windows rendering that code. (See [Na18,NL18] for the role of semantic annotations on educational resources in general.)

```
1  <div about="./Global_OER_Logo.svg.png">
2  <img src="./Global_OER_Logo.svg.png" alt="The Global OER
   Logo" />
3  <span property="dc:title">The Global OER Logo</span>
4  by <a rel="cc:attributionURL dc:creator"
   href="https://commons.wikimedia.org/wiki/User:Jonathasmello
   " property="cc:attributionName">Jonathasmello</a>
5  under <a rel="license"
   href="https://creativecommons.org/licenses/by/3.0/">CC BY
   3.0</a>;
6  from <a rel="dc:source"
   href="https://commons.wikimedia.org/wiki/File:Global_Open_E
   ducational_Resources_Logo.svg">Wikimedia</a>
7  </div>
```

Fig. 1: HTML code for figure with RDFa licensing attribution generated by *emacs-reveal* (simplified excerpt; selected lines explained subsequently)

The *Ljubljana OER Action Plan 2017* [Un17] identifies challenges and proposes actions to mainstream OER. The first challenge, "Building the capacity of users to find, re-use, create and share OER" is to be addressed by ten actions ((a) – (j)), among which action (c) reads as follows:

> *"Disseminate the findings of research on OER to support models of good practice with a focus on cost-effectiveness, sustainability, exploration of new tools and technologies for the creation and sharing of OER;"*

The present paper answers the following research question to address that first challenge in view of action (c): *What attribution requirements are imposed by typical OER licenses and how can they be met in a tool-supported manner that avoids manual efforts and mistakes?* To make that question more precise, in Section 2, OER licensing with Creative Commons is reviewed, pointing out shortcomings related to best practices advertised by

Creative Commons and to the use of CC REL in practice. An attribution approach addressing these shortcomings, which is implemented in the software *emacs-reveal*, is sketched in Section 0. As such, a *usable* practice for the creation and sharing of OER with a *sufficient* amount of machine-readable licensing metadata is advocated.

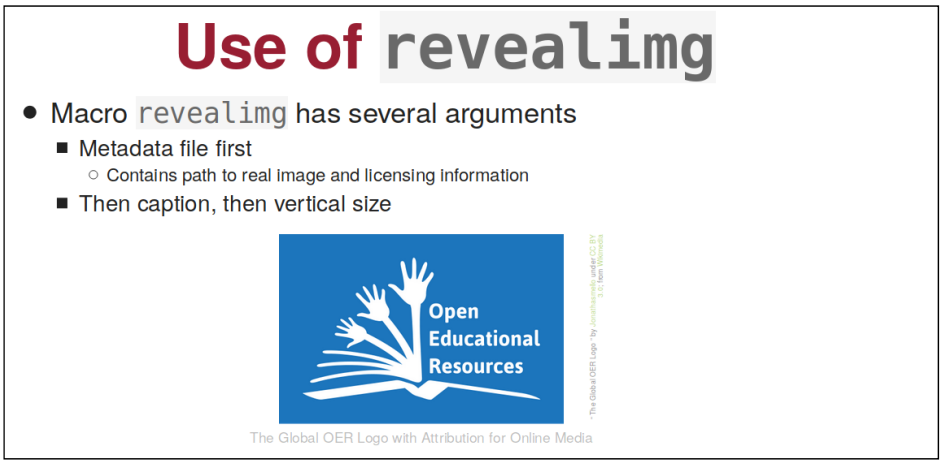



Fig. 2: Sample slide with figure and caption; see Fig. 1 or Fig. 5 for licensing information. (Note that in contrast to the small font size in this screenshot, licensing information to the right of the logo is readable when viewing this slide in a web browser; similarly for Fig. 4.)

## 2 Background

### 2.1 Licenses for OER

As mentioned in the Introduction, OER need to be published under licenses that grant certain freedoms to (re-) users, in particular the 5 Rs of openness, for which typically different forms of Creative Commons (CC) licenses are applied. Among the CC licenses only "CC0" (for resources released into the public domain, without any restrictions on future use), "CC BY" (requiring attribution of the creator) and "CC BY-SA" (requiring attribution and redistribution under same license, "ShareAlike") are "free" licenses according to Creative Commons' licensing considerations [CC19] and "open" licenses as defined by Open Knowledge International [Op19]. Consequently, only those are recommended for OER in [ES18], and only those are considered in the following in their most recent international versions as summarized in Tab. 1. (The subsequent analysis could be extended for any set of licenses, though.)

| Short Name | Full Name | Requirements for use |
|---|---|---|
| CC0 | CC0 1.0 Universal (CC0 1.0) Public Domain Dedication[2] | None |
| CC BY | Attribution 4.0 International (CC BY 4.0)[3] | Attribution to give credit |
| CC BY-SA | Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)[4] | Attribution and re-use with same license |

Tab. 1: Free and open CC licenses for OER

### 2.2 Practical observations

CC licenses are *copyright* licenses, which do not grant patent, trademark, publicity, privacy, or moral rights.[5] For example, if a slideshow under any CC license contains a logo of the publishing organization (say, on its title slide), the rights granted by the CC license to reuse and revise the slideshow and its embedded contents do *not* necessarily extend to the logo, which may be a (trade-) mark that forbids modifications and restricts usage contexts.

Furthermore, a slideshow under one CC license, called primary license subsequently, may reuse OER under different licenses, each of which may need proper attribution. For such compilations, Creative Commons suggests wording such as "Except otherwise noted this work is licensed under […]"[6] to indicate the primary license of the work. Then, only embedded resources with deviating attribution requirements need additional attribution. Given such a situation, an embedded resource without license notice could fall under three different cases:

1. The resource may have been created by the compilation's author, to be covered by the primary license. This should be the typical case.
2. The resource may not be affected by the primary license but come with its own rights and restrictions (e.g., trademarks). Users need to figure this out on their own.
3. The resource may not need attribution if it was released into the public domain (e.g., via CC0). Although the resource's creator does not require attribution, the compilation's author should still indicate the resource's source and license (a) to avoid plagiarism and (b) to enable others to benefit from public domain rights. Hence, authors should avoid this case.

[2] https://creativecommons.org/publicdomain/zero/1.0/
[3] https://creativecommons.org/licenses/by/4.0/
[4] https://creativecommons.org/licenses/by-sa/4.0/
[5] https://creativecommons.org/faq/#rights-other-than-copyright
[6] https://wiki.creativecommons.org/index.php?title=Marking_your_work_with_a_CC_license&oldid=116977

### 2.3 Attribution challenges

When sharing (possibly modified) OER, the two CC licenses of Tab. 1 that require attribution, namely CC BY and CC BY-SA, come with the same set of requirements, abbreviated from their license texts' respective "Section 3(a)(1)":

- Retain if supplied with the OER (a) identification of creator(s), (b) copyright notice, (c) license notice, (d) notice for disclaimer of warranties, (e) hyperlink to OER
- Indicate (previous) modifications
- Indicate license, with text or hyperlink

Notably, a title of the OER is not required with the license versions listed in Tab. 1. Nevertheless, prior versions of those licenses differ in this regard, which lead to the common practice of including the title as well, e.g., under the acronym TASL that is advertised as best practice by Creative Commons [CC18]:

- Title: Include title, if provided
- Author: Name author(s); if specified, use given information
- Source: Include hyperlink to source
- License: Name license and provide hyperlink

Compared to the requirements of "Section 3(a)(1)" listed above, TASL is clearly incomplete as it neither instructs licensees to reproduce notices for copyright and disclaimers nor indications of modifications. Actually, none of the sample attributions advocated as best practices in [CC18] are concerned with these pieces of information. This finding is summarized in the first two columns of Tab. 2.

Finally, note that even the "simple" requirements to include the title or to name the author can be surprisingly difficult in practice. E.g., the OER logo embedded in Fig. 2 was downloaded from Wikimedia Commons[7], which names "Jonathasmello" as author but does not specify a title for the logo. Now, the Wikimedia page links to a UNESCO Web site, where a logo manual[8] can be found. The manual in turn specifies a title and an author. However, the author named in the manual is "2012, Jonathas Mello", contradicting the information on Wikimedia Commons (also the attribution URLs specified at both places disagree). Hence, this logo, which represents the essence of OER, serves as example for incompleteness and inconsistency that is to be expected for manually shared, textual licensing information. Standards for machine-readable metadata promise to reduce such quality issues.

[7] https://commons.wikimedia.org/wiki/File:Global_Open_Educational_Resources_Logo.svg
[8] https://en.unesco.org/sites/default/files/global_oer_logo_manual_en.pdf

### 2.4 Semantic annotations with RDFa

The goal of the Semantic Web is to make Web contents machine-readable by relying on semantic annotations based on standard vocabularies and ontologies. The Resource Description Framework (RDF) is the major language to express knowledge about resources (entities) in the form of so-called RDF triples, consisting of subject, predicate, and object [Ho08]. In our context, a specific OER might be the subject of some triple, its predicate might indicate a creator relationship, and its object might be the creator.

RDF triples can be represented (serialized) in different forms, and RDFa (Resource Description Framework in Attributes) [He15] specifies how to embed RDF triples in HTML attributes. Examples for machine-readable Creative Commons licensing with RDFa in HTML documents are given in [CC14] and in Fig. 1. E.g., line 6 in Fig. 1 demonstrates the use of RDFa to annotate a usual hyperlink with semantic information in the `rel` attribute, indicating that the hyperlink identifies the source of some figure.

### 2.5 CC REL

CC REL (The Creative Commons Rights Expression Language [Ab12]) is a standard for license terminology based on RDF (with HTML embedding via RDFa). For example, line 4 of Fig. 1 shows how a hyperlink can be used to attribute an author with name and hyperlink.

The vocabulary offered by CC REL to indicate license and attribution information (the so-called "Work properties" of CC REL) is listed in the third column of Tab. 2. Apparently, CC REL enables OER *authors* as licensors to specify licensing information according to TASL, but is not sufficient to express all attribution requirements that could be demanded when licensees (re-) *use* OER. In particular, CC REL lacks means to indicate modifications of OER.

For example, suppose that Alice publishes a figure with attribution license. Bob improves that figure and publishes his improved version with a different title at a different place, possibly under a different (compatible) license that also requires attribution. The attributes offered by CC REL could either record license information provided by Alice or by Bob. As Bob has created a new work, he uses the CC REL attributes for his information, leaving no place to record Alice's information, or, more generally, a description of the figure's *provenance*. In lack of a standard, this gap is addressed in a pragmatic way with *emacs-reveal* as explained subsequently.

## 3 Attribution in *emacs-reveal*

The software *emacs-reveal* supports the creation of OER slideshows (HTML based on the presentation framework *reveal.js*[9] with audio explanations) and meets the following technical requirements (subsuming the ALMS framework of [Hi10] as argued in [Le19]):

- OER should be usable (for learning) with FLOSS on (almost) any device, also mobile and offline.
- OER should be editable with FLOSS (this requires source file access).
- OER should be re-usable under the Single Sourcing paradigm (see [Roc01]), which enables reuse and revision from a single, consistent source without copy&paste (copy&paste creates isolated copies, where the reconciliation of changes and improvements by different individuals would be almost impossible).
- OER should offer a separation of contents from layout (then, experts for content do not need to be design experts as well; also, cross-organizational collaboration is supported where each organization can apply its own design guidelines).

| Requirement | TASL | CC REL | *emacs-reveal* |
|---|---|---|---|
| Name creator(s) | A(uthor) | cc:attributionName, cc:attributionURL | cc:attributionName, cc:attributionURL |
| Reproduce copyright notice | - | - | copyright |
| Reproduce license notice | L(icense) | license | licenseurl, licensetext |
| Reproduce disclaimer of warranties | - | - | copyright, permit |
| Include hyperlink to OER | S(ource) | dc:source | dc:source, sourcetext |
| Indicate modifications | - | - | imgadapted |
| Indicate license | L(icense) | license | licenseurl, licensetext |
| - | T(itle) | dc:title | dc:title, imgalt |
| - | - | cc:morePermissions | permit |

Tab. 2: Comparison of CC attribution requirements and mechanisms

[9] https://revealjs.com/

- OER should be defined in a lightweight markup language, which is easy to learn and which enables the use of industrial-strength version control systems such as Git for the management of OER collaboration (comparison, revision, merge).

Slideshows are defined in Org mode text files, where figures can be embedded using two macros `revealimg` and `reveallicense` as indicated in Fig. 3, whose code produces two slides, excerpts of which are shown in Fig. 2 and Fig. 4 (both display an OER logo). Both macros produce figures where licensing information is rotated to the right with the default CSS styling of *emacs-reveal*. With `revealimg`, the figure is horizontally centered and has a caption, while `reveallicense` embeds the figure in the surrounding paragraph or list (without caption). The first argument to both macros is the name of a file with metadata, here `"Global_OER_Logo.svg.meta"`, whose contents are reproduced in Fig. 5.

```
* Use of ~revealimg~
  - Macro ~revealimg~ has several arguments
    - Metadata file first
      - Contains path to real image and licensing information
    - Then caption, then vertical size

  {{{revealimg("Global_OER_Logo.svg.meta","The Global OER Logo
with Attribution for Online Media","30vh")}}}

* Use of ~reveallicense~
  - Here, the logo appears in reduced size without caption
    {{{reveallicense("Global_OER_Logo.svg.meta","15vh")}}}
  - As part of a simple list
```

Fig. 3: Sample Org mode code to create two slides

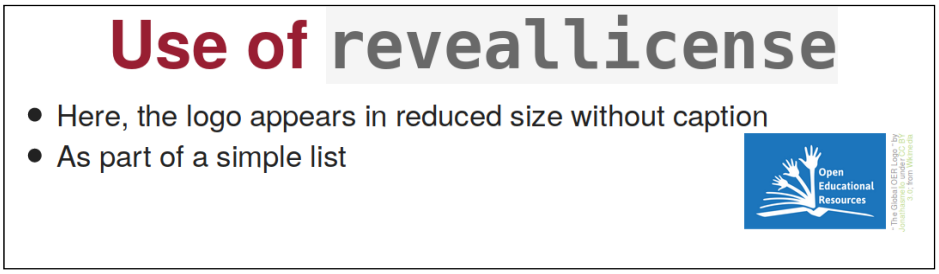



Fig. 4: Sample slide with figure without caption (see Fig. 1 or Fig. 5 for licensing information)

```
1   ; Semicolon starts comment until end of line (Emacs Lisp).
2   ; Lines 13-16 occur in comments; they illustrate available
    options.
3
4   ((filename . "./Global_OER_Logo.svg.png")
5    (licenseurl .
    "https://creativecommons.org/licenses/by/3.0/")
6    (licensetext . "CC BY 3.0")
7    (cc:attributionName . "Jonathasmello")
8    (cc:attributionURL .
    "https://commons.wikimedia.org/wiki/User:Jonathasmello")
9    (dc:source .
    "https://commons.wikimedia.org/wiki/File:Global_Open_Educat
    ional_Resources_Logo.svg")
10   (sourcetext . "Wikimedia")
11   (dc:title . "The Global OER Logo")
12   (texwidth . 0.2)
13  ; (imgalt . "If the title is not suitable as alt text.")
14  ; (imgadapted . "Indicate modifications/provenance.")
15  ; (permit . "Indicate special permissions/disclaimers.")
16  ; (copyright . "Reproduce copyright notice of source.")
17  )
```

Fig. 5: Sample metadata with license information

Note that the figures containing the OER logo are just screenshots in this document. In the corresponding slideshow, attributed author, source, and license name are hyperlinked as required by the license. Thus, in contrast to the slideshow, Fig. 2 and Fig. 4 are not self-contained and cannot be distributed without additional licensing information. For the purposes of this document, necessary information is shown in Fig. 1 as well as in Fig. 5.

The different metadata attributes of *emacs-reveal* occurring in Fig. 5 are also sorted into column 4 of Tab. 2. Most extensions over CC REL have been created for pragmatic reasons: While CC REL can indicate a `license` URL for use in a hyperlink, in practice the hyperlink also needs a text, for which CC attribution requires reproduction; hence, *emacs-reveal* uses two attributes, `licenseurl` and `licensetext`. Similarly, a `sourcetext` is added as text for the `dc:source` hyperlink. As images in HTML should provide an `alt` attribute with a textual description for accessibility reasons, `imgalt` can be used to specify this information. In addition, modifications can be recorded with `imgadapted` (for provenance information as mentioned in Section 2.5), while `permit` and `copyright` allow licensors to document and licensees to reproduce additional licensing information. Actually, until now the author has not seen a need for `cc:morePermissions`, which is why this attribute is not part of *emacs-reveal* (adding it is left for future work). In addition, `copyright` has only been added to meet CC attribution requirements, but has not been used so far. Besides, `texwidth` hints at the fact that Org slideshows can also be exported via LaTeX to PDF; then width specifications for figures improve the layout.

The macros `revealimg` and `reveallicense` come with additional parameters that influence the result. In particular, license information can be displayed in short form, which is useful to cover case (3) of Section 2.2, namely a brief note documenting CC0 (public domain) licensing for a figure, or omitted entirely to cover case (1) of Section 2.2, namely figures whose licensing terms agree with those of the overall slideshow.

Finally, readers might wonder about the choice of the format for metadata files (such as the one shown in Fig. 5). A plain text file with a list data structure in Emacs Lisp was chosen for the simple reason that the bulk of *emacs-reveal* is running inside Emacs anyways, where such lists can be parsed easily. Besides, such simple text files can be created and edited by a wide range of users with any text editor. Regardless of which format might be chosen by other tools or future standards, essentially key-value pairs need to be recorded. Clearly, a variety of formats could be used, in particular RDF (in one of its serialization formats) or plain JSON. Whenever other tools would benefit from a different format, it should not be difficult to support such formats in *emacs-reveal* as well.

As of March 18, 2019, a public GitLab repository[10], which the author created in March 2018 to share class material and talks, hosts 148 metadata files out of which 125 (attributed to 19 different creators) carry free and open licenses as shown in Tab. 3. Thanks to the sketched approach, the author does not need to track the various creators and licenses manually. New contributors to the repository are welcome. When re-using figures from that repository via *emacs-reveal*, the software accesses machine-readable, structured metadata. Thus, quality issues illustrated in Section 2.3 (which result from human interpretation and manual copying of textual licensing annotations) are avoided.

| License | Usage count |
|---|---|
| Public domain | 5 |
| CC 1.0 | 49 |
| CC BY 2.0 | 1 |
| CC BY 2.5 Australia | 4 |
| CC BY 3.0 | 2 |
| CC BY-SA 2.0 | 1 |
| CC BY-SA 2.5 | 2 |
| CC BY-SA 3.0 | 23 |
| CC BY-SA 4.0 | 37 |
| GPLv2 | 1 |

Tab. 3: Breakdown of metadata files on public repository by OER license

[10] https://gitlab.com/oer/figures

## 4 Conclusions

Proper licensing when creating and reusing OER is a tedious task, if necessary metadata needs to be collected and applied manually. As argued in Section 2, metadata available in different places may be incomplete and inconsistent, and the standard vocabulary CC REL does not meet all requirements for the reuse of OER. Hence, the following research question arises: *What attribution requirements are imposed by typical OER licenses and how can they be met in a tool-supported manner that avoids manual efforts and mistakes?*

To answer this question, an approach towards creation and sharing of metadata was presented in Section 3, which is implemented in *emacs-reveal*. Here, simple text files store metadata describing figures, and metadata is shared along with figures. When embedding such a figure into a slideshow, license information is displayed automatically. Importantly, license metadata is both machine-readable (based on RDFa) and human-readable. Thus, all RDFa compatible tools (e.g., search engines or browser extensions such as OpenLink Structured Data Sniffer[11]) can extract license information even if they are unaware of *emacs-reveal*.

The metadata vocabulary of *emacs-reveal* builds upon CC REL (summarized in Tab. 2), but has been extended to meet real-life requirements in pragmatic ways. Clearly, a need for further standardization exists, which is the subject of future work.

Finally, the need for sharing *two* files (image and metadata) requires some discipline which would not be necessary if metadata was embedded directly, e.g., with a standard such as XMP[12]. Until such standards are more widespread, *emacs-reveal* offers a novel approach to gather experiences with OER and machine-readable license attribution in practice.

Bibliography

[Ab12] Abelson, H.; Adida, B.; Linksvayer, M.; Yergler, N.: CC REL: The Creative Commons Rights Expression Language. In (Dulong De Rosnay, M.; De Martin, J.C., ed.): The Digital Public Domain: Foundations for an Open Culture. Open Book Publishers, 2012.

[AS16] Amiel, T.; Soares, T.: Identifying Tensions in the Use of Open Licenses in OER Repositories. The International Review of Research in Open and Distributed Learning 17 (3), 2016.

[CC14] Creative Commons: Marking Works Technical. Wiki, 2014. https://wiki.creativecommons.org/index.php?title=Marking_Works_Technical&oldid=99270, accessed 17/03/2019.

[CC18] Creative Commons: Best Practices for Attribution. Wiki, 2018. https://wiki.creativecommons.org/index.php?title=Best_practices_for_attribution&oldid=116817, accessed 17/03/2019.

[11] http://osds.openlinksw.com/
[12] https://www.iso.org/standard/57421.html

[CC19] Creative Commons: Licensing Considerations. 2019. https://creativecommons.org/share-your-work/licensing-considerations/, accessed 17/03/2019.

[Cr08] Crowston, K.; Wei, K.; Howison, J.; Wiggins, A.: Free/Libre Open-Source Software Development: What We Know and What We Do Not Know. ACM Comput. Surv. 44 (2): 7:1–7:35, 2008.

[ES18] Ebner, M.; Schön, S.: |. In: Miglbauer, M.; Kieberl, L.; Schmid, S. (eds.). Hochschule digital.innovativ | #digiPH. Tagungsband zur 1. Online-Tagung, Graz: FNMA, S. 183–194.

[Fa16] Falconer, I. J.; Littlejohn, A.; McGill, L.; Beetham, H.: Motives and Tensions in the Release of Open Educational Resources: The UKOER Program. Australasian Journal of Educational Technology 32 (4), 2016.

[Iv15] Ivan, H.; Adida, B.; Sporny, M.; Birbeck, M.:RDFa 1.1 Primer - Third Edition. 2015. https://www.w3.org/TR/rdfa-primer/, accessed 17/03/2019.

[Hi10] Hilton, J.; Wiley, D.; Stein, J.; Johnson, A.: The four 'R's of openness and ALMS analysis: frameworks for open educational resources. Open Learning 25 (1): 37–44, 2010.

[Ho08] Horrocks, Ian: Ontologies and the Semantic Web. Commun. ACM 51 (12): 58–67, 2008.

[Op19] Open Knowledge International: Conformant Licenses. 2019. https://opendefinition.org/licenses/, accessed 17/03/2019.

[Le19] Lechtenbörger, J.: Erstellung und Weiterentwicklung von Open Educational Resources im Selbstversuch. *MedienPädagogik* 34: 101–17, 2019.

[Na18] Nahhas, S.; Bamasag, O.; Khemakhem, M.; Bajnaid, N.: Added Values of Linked Data in Education: A Survey and Roadmap. *Computers* 7 (3), 2018.

[NL18] Navarrete, R.; Luján-Mora, S.: A Quantitative Analysis of the Use of Microdata for Semantic Annotations on Educational Resources. *Journal of Web Engineering* 17 (1&2): 045–072, 2018.

[Ro01] Rockley, A.: The Impact of Single Sourcing and Technology. Technical Communication 48 (2). Society for Technical Communication: 189–93, 2001.

[SD11] Schulte, E.; Davison, D.: Active Documents with Org-Mode. *Computing in Science Engineering* 13 (3): 66–73, 2011.

[Ro01] Rockley, A.: The Impact of Single Sourcing and Technology. Technical Communication 48 (2). Society for Technical Communication: 189–193, 2001.

[Un12] UNESCO: 2012 Paris OER Declaration. https://en.unesco.org/oer/paris-declaration, accessed 17/03/2019.

[Un17] UNESCO: Ljubljana OE Action Plan 2017. https://en.unesco.org/sites/default/files/ljubljana_oer_action_plan_2017.pdf, accessed 17/03/2019.

[Wi14] Wiley, D.: The Access Compromise and the 5th R. 2014. https://opencontent.org/blog/archives/3221, accessed 17/03/2019.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 217

# Praktische Erfahrungen mit einem elektronischen Prüfungsprozess-Management für eAssessments

Marc Sauer[1], Manuel Froitzheim[2] und Andreas Hoffmann[3]

**Abstract:** Die Durchführung von reliablen, validen und rechtssicheren Prüfungen in elektronischer Form wurde praktisch getestet und inzwischen verstetigt an der Universität Siegen. Die Erfahrungen und die Struktur des Prüfungsprozess-Managements werden in diesem Praxisbeitrag dargestellt.

**Keywords:** eAssessment, Prozessmanagement, Organisationsmanagement, Qualitätssicherung

## 1 Einleitung

E-Prüfungen ermöglichen eine schnelle Durchführung und Korrektur von Prüfungen. Sie bieten aber auch Chancen für ein strukturiertes Organisations- und Prozessmanagement und neue Wege der Qualitätssicherung (z.B. inhaltliche Reviews, Kriteriengetriebene Prüfungszusammenstellung, faire Post-Prüfungs- und Einspruchsprozesse, neue Statistiken und Aufbau einer hochschulweiten Datenbank) [Eh13]. In Anlehnung an die Beschreibung typischer Prozesse für schriftliche Klausuren von Schmees und Horn werden in diesem Beitrag Prozesse betrachtet, die zu reliablen, validen und rechtssicheren (E-)Prüfungen führen und somit das Qualitätsmanagement sowie die Organisation und Kommunikation im Prüfungsprozess steuern und optimieren können [SH14]. Abbildung 1 zeigt dabei ein häufig vorzufindendes Prozessmanagement bei E-Prüfungen, wie es auch in Softwareprodukten zur Prüfungsadministration und Durchführung integriert ist (hier am konkreten Beispiel der Universität Siegen, Schmees und Horn skizzieren jedoch einen ähnlichen Ablauf der Universität Bremen [SH14]).

[1] Universität Siegen, Lehrstuhl für Betriebssysteme und verteilte Systeme, Hölderlinstraße 3, 57068 Siegen, marc.sauer@uni-siegen.de
[2] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57068 Siegen, froitzheim@zoebis.de
[3] Universität Siegen, Lehrstuhl für Betriebssysteme und verteilte Systeme, Hölderlinstraße 3, 57068 Siegen, andreas.hoffmann@uni-siegen.de

doi:10.18420/delfi2019_344

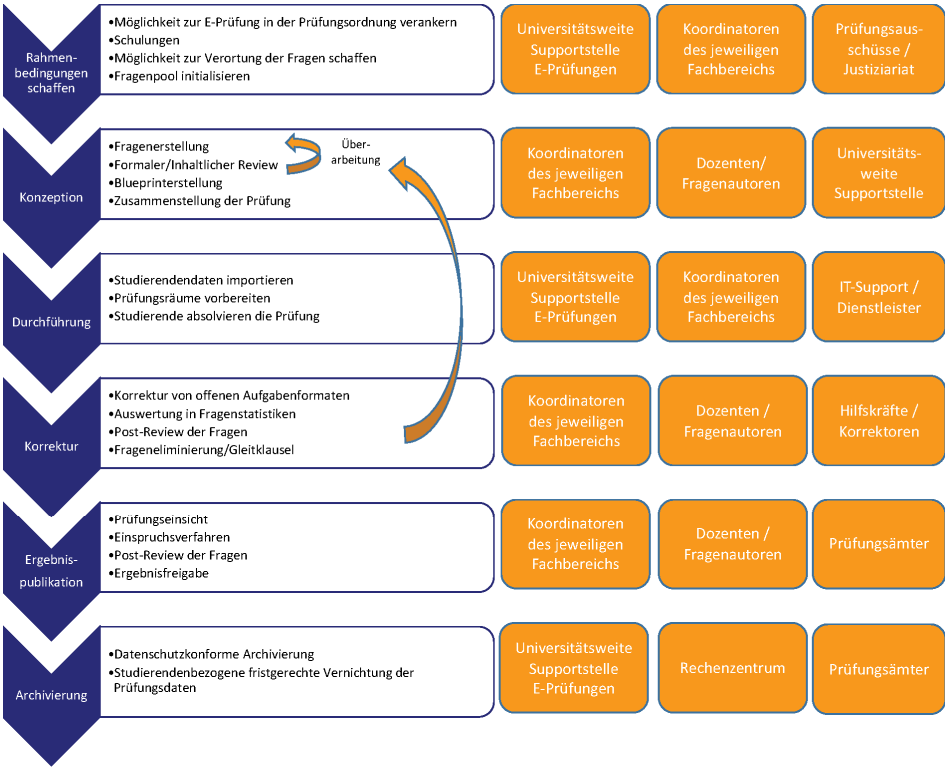



Abb. 1 Prozessmanagement E-Prüfungen (Universität Siegen)

## 2 Ganzheitliche Betrachtung des Prüfungsprozesses

Durch die ganzheitliche Betrachtung des Prüfungsprozesses und dessen digitaler und organisatorischen Abbildung an der Universität Siegen, konnte der gesamte Prozess sowohl effizienter als auch effektiver gestaltet werden. Im gesamten Prüfungsprozess wird kein Blatt-Papier benötigt und alle Schritte können digital durchgeführt werden.

### 2.1 Erstellung der Prüfung

Bereits die Erstellung der Prüfung und der kollaborativen Möglichkeit Aufgaben zu einer Veranstaltung zu erstellen und dessen formale und inhaltliche Korrektheit mit einem standardisierten Verfahren zu überprüfen, vermeidet die typischen Probleme (Versionskonflikte, unterschiedliche Formatierung, doppelte Aufgaben, unverschlüsselte Übertragung per E-Mail). Vogt und Schneider definieren allgemeine Anforderungen an ein Autorenwerkzeug und die Fragenverwaltung für elektronische Prüfungen [VS09].

Dementsprechend können die Prüferinnen und Prüfer in einer gesicherten webbasierten Umgebung (an der Universität Siegen wird das Tool Q-Exam® Institution verwendet) vor der Prüfung die unterschiedlichsten Fragentypen anlegen. Zur Auswahl stehen sowohl Fragentypen mit geschlossenen Aufgabenformaten, wie zum Beispiel Single Choice, Multiple Choice oder Lückentexte, als auch Fragentypen mit offenen Antwortformen, wie zum Beispiel Freitextaufgaben. Alle Fragen werden in einem öffentlichen Pool für die Lehrenden der Lehrveranstaltung gespeichert, nachdem jede Frage einen strukturierten Review Prozess durchlaufen hat. Für die Speicherung wird jede Frage verortet bezüglich der Fakultät, dem Lehrstuhl und der Lehrveranstaltung. Später können bei der Fragenzuordnung zu einer Prüfung geeignete Fragen besser gefunden werden, wenn auch Wissensdimensionen, Prozessdimensionen und Lehrziele angegeben wurden. Durch die detaillierte Verortung können Vorgaben, dass zum Beispiel 1/3 der Fragen zu einem bestimmten Teil der Vorlesung ausgewählt werden soll, auch maschinell überprüft werden.

Der Review-Prozess für jede Frage sieht vor, dass die Frage auf formale und inhaltliche Kriterien überprüft wird. Der formale Review überprüft zum Beispiel die Fragenformulierung und den von der Autorin oder dem Autor ausgewählten Fragetypen. Zum Beispiel sind Fragen mit doppelter Verneinung in der Fragestellung nicht zielführend in einer Prüfung. Neben dem formalen Review wird von einer anderen Person eine inhaltliche Überprüfung durchgeführt. Die Person überprüft die Fragestellung, ggf. die Antwortmöglichkeiten und den Erwartungshorizont. Werden im formalen und fachlichen Review Fehler entdeckt, können diese dokumentiert werden. Die Autorin oder der Autor der Frage wird anschließend automatisiert per E-Mail informiert, dass eine Frage nicht die Gütekriterien für eine gute Frage erfüllt. Nach der Überarbeitung wird die Frage wieder in den Review aufgenommen und erneut überprüft.

Basierend auf der Datenbank der Fragen wird eine Prüfung zusammengestellt. Bei der Zusammenstellung der Prüfung können formale Kriterien angegeben werden, damit diese automatisiert kontrolliert werden. Zum Beispiel der Anteil an geschlossenen und offenen Aufgaben oder auch die Anzahl an Fragen insgesamt.

## 2.2 Durchführung von E-Prüfungen

Die Vorteile der eigentlichen Durchführung von E-Prüfungen wurden bereits umfangreich diskutiert [HS17a]. Durch die ganzheitliche Betrachtung des Prüfungsprozesses ergeben sich auch bei der Durchführung weitere Vorteile. Die Prüfungsteilnehmerinnen und Prüfungsteilnehmer werden aus dem Campus Management System (an der Universität Siegen wird HISinOne verwendet) im Vorfeld in die Prüfungsumgebung importiert. Dadurch kann zu Beginn der Prüfung eine automatisierte Überprüfung der Prüfungsberechtigung erfolgen. Ist eine Studentin oder ein Student nicht zur Prüfung angemeldet, erfolgt keine Freigabe der Prüfung und eine Teilnahme wird technisch ausgeschlossen. Dies ermöglicht auch die parallele Durchführung unterschiedlicher Prüfungen in einem Raum. Die Bearbeitungszeit bei den parallel stattfindenden Prüfungen

muss identisch sein, damit keine Störungen durch umherlaufende Studierende entstehen. Ansonsten erhalten die Studierenden jeweils nur die eigene Prüfung. Mit einer alternierenden Sitzordnung ist auch eine engere Bestuhlung möglich, weil ein Abschreiben bei der Nachbarin oder dem Nachbarn nicht möglich ist.

### 2.3 Kollaborativer Korrekturprozess

Im kollaborativen Korrekturprozess, insbesondere bei Freitextaufgaben sind die maschinengeschriebenen Texte der Studierenden schneller zu lesen als handgeschriebene Texte [HS17b]. Die Korrektur kann auch ortsunabhängig über ein gesichertes virtuelles privates Netzwerk (VPN) erfolgen. Dies ist insbesondere in der vorlesungsfreien Zeit, in der die Prüferinnen und Prüfer sehr oft im Ausland an Konferenzen teilnehmen von Bedeutung.

Zur Dokumentation der Korrektur können Anmerkungen eingegeben werden und durch die Studierenden in der Klausureinsicht gelesen werden, welche nach der Veröffentlichung der Ergebnisse stattfindet. Die bessere Dokumentation der Korrektur führt zu weniger Einsprüchen von Studierenden. In der elektronischen Prüfungseinsicht (derzeit unter Aufsicht an den Rechnern der Prüferinnen und Prüfer) können die Studierenden zu den einzelnen Fragen Kommentare bzw. Anträge auf nachträgliche Stellungnahme und ggf. Nachkorrektur abgeben, die im Anschluss an die Frist zur Einsichtnahme von den Prüferinnen und Prüfer oder ggf. auch von den Autorinnen und Autoren geprüft werden. Im ersten Prüfungstermin des Wintersemesters 2018/19 wurden zu 1833 Prüfungen nur drei Anträge auf eine Punkteänderung gestellt. Für alle anderen Studierenden waren die Anmerkungen ausreichend, um die Bewertung nachzuvollziehen. Im Wintersemester 2017/18 gab es circa 25 Anträge auf eine Punktekorrektur bei gleich vielen durchgeführten papierbasierten Prüfungen.

Nach Abschluss der Korrektur wird jede Aufgabe mit statistischen Kennzahlen, zum Beispiel mit dem Maß der Trennschärfe, überprüft. Daraufhin wird die Frage gegebenenfalls wieder an die Autorin oder den Autor zur erneuten Überarbeitung zurückgegeben, damit die Frage bei einer erneuten Nutzung in einem der Folgeprüfung in korrigierter Version genutzt werden kann. Ebenso wird entschieden, ob die Bewertung einzelner Fragen (z.B. Frage ersatzlos aus der Prüfung nehmen) zu einem korrigierten Gesamtergebnis der Prüfung führt. Dies führt gegebenenfalls zu einem erneuten publizieren der Ergebnisse (vgl. Abbildung 2).

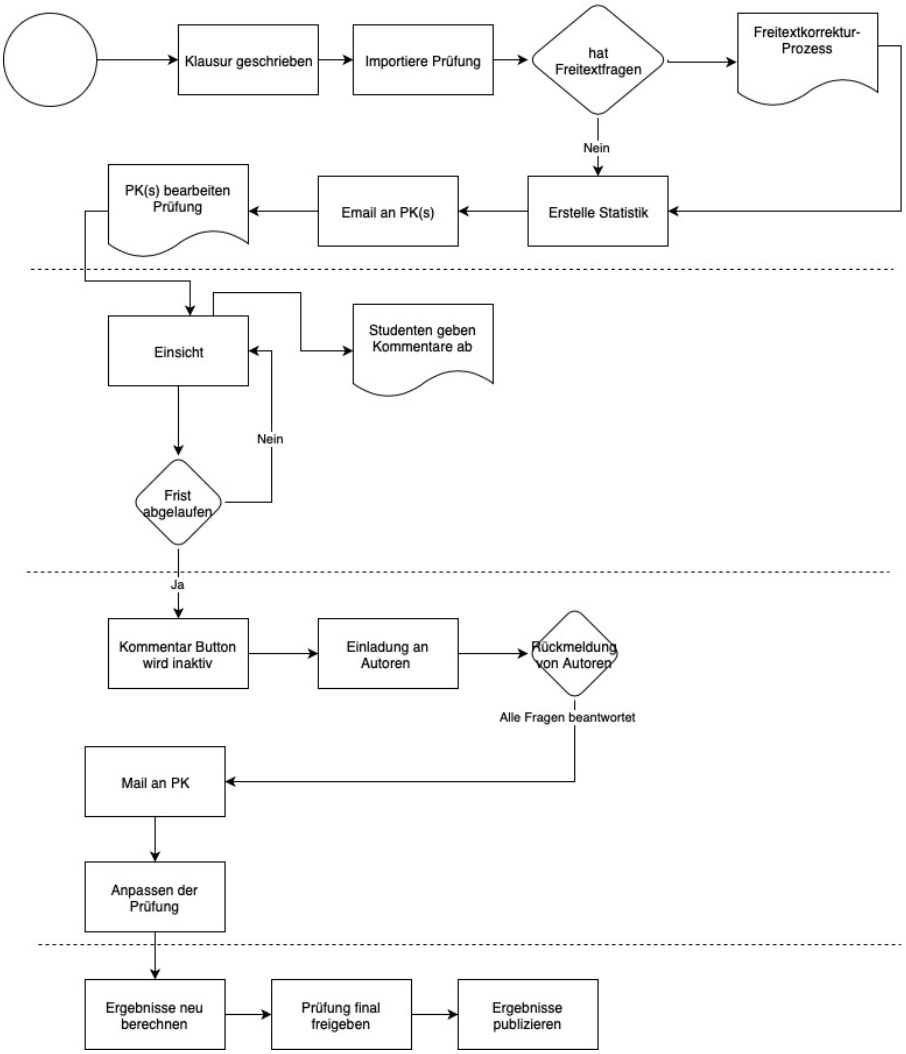



Abb. 2 Korrekturprozess der Prüfung (spezifisch für die Universität Siegen, PK = Prüfungskoordinator)

## 2.4 Archivierung der Prüfungen

Zum Abschluss des ganzheitlichen Prozesses ist die Archivierung der Prüfungen einfacher als bei Papier-Klausuren. In den Lehramtsstudiengängen ist an der Universität Siegen das jeweilige Institut für die Archivierung der Prüfungen verantwortlich und dies führt bei zum Beispiel circa 10.000 Prüfungen pro Jahr an einem einzigen Institut zu einem

umfangreichen Bestand an Prüfungen. Diese Prüfungen müssen entsprechend der rechtlichen Vorgaben zur Archivierung der jeweiligen Prüfungsordnung archiviert und geordnet werden. Die Archivierung der elektronischen Prüfungsdaten erfolgt an der Universität Siegen auf universitätseigenen Servern, die folglich den rechtlichen Vorgaben der Datensicherung und Datenhaltung genügen. Insbesondere ist es durch automatische Zeitstempel viel einfacher möglich, die nach Ende der Aufbewahrungsfrist (an der Universität Siegen bspw. drei Jahre nach endgültigem Bestehen oder nicht Bestehen der Prüfungsleistung) zu löschenden Prüfungsdaten ausfindig zu machen. Perspektivisch könnte hier auch eine automatische Löschung der Prüfungsdaten nach der vorgegebenen Aufbewahrungsfrist umgesetzt werden.

## 3 Fazit

Das Prüfungsprozess-Management für eAssessments hat dazu geführt, dass der gesamte Prüfungsprozess abgebildet wurde und die Abläufe strukturiert stattfinden. Dies führt zu einer deutlichen Erleichterung für alle Beteiligten, weil viele Aufgaben orts- und zeitunabhängig stattfinden können.

Literaturverzeichnis

[Eh13] Ehlers et al.: Prüfen mit Computer und Internet. Didaktik, Methodik und Organisation von E-Assessment. In: Ebner, M.; Schön, S. (Hrsg.): Kompetent Prüfungen gestalten. 53 Prüfungsformate für die Hochschullehre, 2. Auflage, Epubli 2013, Berlin, online. 2013.

[HS17a] Hoffmann, A.; Sauer, M.: E-Klausur. In: J. Gerick, A. Sommer & G. Zimmermann (Hrsg.): Kompetent Prüfungen gestalten. 53 Prüfungsformate für die Hochschullehre, Waxmann 2017, S. 46ff. 2017.

[HS17b] Hoffmann, A.; Sauer, M.: Juristische Volltext-E-Klausuren. Tagungsbroschüre E-Prüfungssymposium. Neue Prüfungsformen im Zeitalter der Digitalisierung, S. 40f., https://e-pruefungs-symposium.de/wp-content/uploads/2017/09/Broschuere_ePS2017.pdf, zuletzt aufgerufen am 25.07.2018

[SH14] Schmees, J.; Horn, J.: E-Assessments an Hochschulen: Ein Überblick: Szenarien. Praxis. E-Klausur-Recht. In: Reihe ELAN e.V. (Hrsg.): Digitale Medien in der Hochschullehre, Band 1, Waxmann Verlag, Münster, 2014.

[VS09] Vogt, M.; Schneider, S.: E-Klausuren an Hochschulen: Didaktik – Technik – Systeme – Recht – Praxis. Koordinationsstelle Multimedia, JLU Gießen, 2009.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 223

# Predictive Algorithms in Learning Analytics and their Fairness

Shirin Riazy[1] and Katharina Simbeck[2]

**Abstract:** Predictions in learning analytics are made to improve tailored educational interventions. However, it has been pointed out that machine learning algorithms might discriminate, depending on different measures of fairness. In this paper, we will demonstrate that predictive models, even given a satisfactory level of accuracy, perform differently across student subgroups, especially for different genders or for students with disabilities.

**Keywords:** Learning Analytics, Fairness, OULAD, At-Risk Prediction

## 1 Introduction

While the issue of fairness has been discussed in policing, operational systems and recommender systems, many machine learning (ML) areas, including learning analytics, still lack such contributions. In what ways can biases in data – possibly unjustifiably – influence algorithms? Since predictions in learning analytics are made to improve tailored educational interventions, the fitness of interventions requires the appropriateness of preceding predictions. On one hand, any inappropriate (discriminatory) intervention may further intensify discrimination [O'17]. On the other, a student requiring more involvement but not receiving it, might be neglected. Therefore, the validity of learning analytics algorithm must be analysed with regards to their predictive accuracy for minorities.

As of now, the definition of fairness in algorithms is highly ambiguous. Many definitions of fairness have been suggested and, as noted by [VR18], the choice of this definition should strongly depend on the context.

In this position paper, we would like to review some of the emerging definitions of fairness and transfer and evaluate their fit in the context of learning analytics, specifically the fairness of score predictions.

### 1.1 Definitions of Fairness

As mentioned briefly, a vast amount of different definitions of fairness have surfaced, serving different purposes and being designed to different use-cases. [VR18] have

[1] Hochschule für Technik und Wirtschaft, Berlin
[2] Hochschule für Technik und Wirtschaft, Berlin

doi:10.18420/delfi2019_305

gathered and explained 20 definitions of fairness, which are partially exclusive [Ch17]. We would like to briefly introduce selected definitions of fairness, which had the largest impact on the farness debate and research, measured by the number of citations. Also, since the research in this area is rapidly evolving, we would like to include one more fairness measure, which was only recently published, but is relevant as it was applied to a learning analytics scenario [GBB19].

The fairness measures rely on accuracy measures for predictive models. First, we would like to introduce some widely used model accuracy measures:

- We abbreviate the classical terms of binary classification in the usual way: true positive (TP), false positive (FP), true negative (TN), false negative (FN).
- The positive predictive value (PPV) is defined as TP / (TP + FP).
- The true positive rate (TPR) is defined as TP / (TP + FN).
- The area under the curve (AUC) of the receiver operator characteristic (ROC) graph, which shows the difference between predicted and real values for different choices in the trade-off between TP and FP rates.

Based on those measures, the model accuracy can be compared between different subgroups. Ideally, models should predict success equally well for subgroups, e.g. for female and male students: Predictive Parity (PP) demands equal PPV between groups [Ch17]. Equalised Odds (EO) demands equal TPR between groups [HPa16]. Slicing Analysis (SA) demands equal AUC between groups [GBB19]. In reality however, it cannot be expected that model accuracy is always consistent across all subsets of data. Therefore, a threshold needs to be defined, which delineates fair from unfair. For PP and EO, we set this threshold to be 0.05, as it was done in [VR18]. Since SA values mostly lie between 0.01 and 0.03 (see [GBB19]), we set the threshold to 0.02.

## 2 Related Work: Performance Prediction of Students and Fairness

[Si13] describes trend analysis and prediction as one of the primary applications of learning analytics. In order to build early warning systems and to identify risk, techniques of modelling users, relationship mining or knowledge domain modelling have been used [Si13]. Though ethical usage of user data was always called for (see for example [GD12, SP13]), this has a particular focus on privacy and data ownership [Si13].

The identification of at-risk students, their performance and attrition risk has been considered in literature in an attempt to find a basis for intervention to enhance student retention [AP12]. In order to build “early warning systems”, data from virtual learning environments has been analysed to make predictions on learner behaviour [MD10]. A wide variety of procedures have been used to predict students at risk of failing or dropping out. These range from statistical methods, such as logistic regression [HZZ17] over naïve Bayes classification [Ku15] to predictions via neural networks [FY15].

# 3 Methodology

The data set that was used in the context of this paper is known as OULAD (Open University Learning Analytics Dataset). The open university (UK) has published this anonymised data set from their virtual learning environment (VLE) in 2017 [KHZ17] and it roughly consists of 32,000 samples, meaning student data from one course and one term. For students taking more than one course, these are counted as multiple samples.

Since we would like to reproduce realistic applications of ML algorithms in learning analytics, we tried to replicate publications, where this exact data set was used, such as in [HZZ17] and [Ku15]. Note, however, that we did not copy the exact method, since e.g. [Ku15] make a weekly prediction of at-risk students. In their paper, they split the features into demographic data (gender, age, prior education, disability) and virtual learning environment (VLE) data (number of assessments, days of access, number of clicks, number of logs). We used the following features to classify the outcome (fail, pass) of the course:

- Demographic: gender, prior education, age range, disability (boolean flag)
- VLE: number of assessments, number of different days on which material was accessed, average number of clicks, total number of logs
- Other: number of other credits studied in selected term, number of previous attempts

In order to avoid overfitting, we split the data set into a training, test and validation set (see e.g. [Mu12]). We chose to split the data according to the terms during which the modules took place. Out of these terms, two took place during the year 2013, which we chose as the training set (sample size: 13529). Two took place during the year 2014, which we chose as test set (sample size: 7804) and validation set (sample size: 11260). We split the terms in a way such that the resulting data sets would not vary much in sample size.

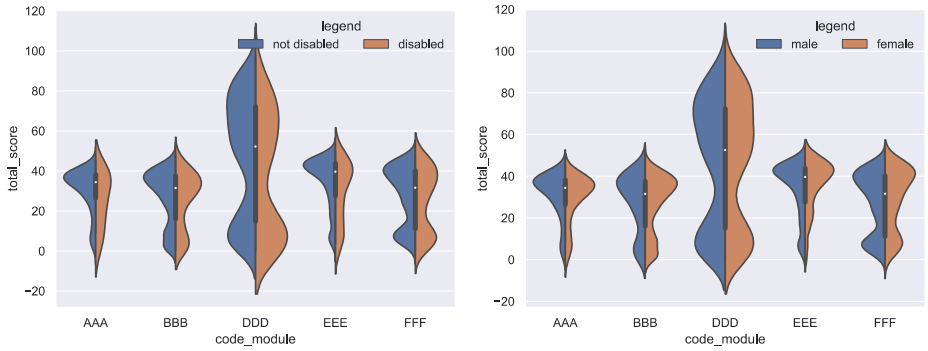



Figure 1: Violin plots showing the distributions of scores of non-disabled/disabled and female/male participants of the courses.

The distributions of the course scores (see Figure 1) show that, when grouping by gender, the distributions of the total scores of the different courses do not vary much. In contrast to this, the distributions greatly vary, when comparing disabled (sample size: 3164) and non-disabled students. In total, 47,2% of students in the dataset passed the course. The rate was higher for female students (48.4%) in comparison to male students (46.2%) and higher for students without disability (48.1%) in comparison to students with disability (38.1%).

### 3.1 Prediction of Course Outcome

Naïve Bayes (NB), k-Nearest-Neighbours (kNN) and Support Vector Machines (SVM) were used in [Ku15] to build a course-outcome prediction model for students. We replicated these algorithms and applied them to the OULAD data set. For more detailed descriptions of these algorithms, we would like to refer the reader to [Mu12]. Our results for the prediction systems are summarized in the following table, where the accuracy is defined as the relative sum of all correct classifications.

| | NB | kNN | SVM |
|---|---|---|---|
| Test Set | 0.728 | 0.760 | 0.776 |
| Validation Set | 0.756 | 0.751 | 0.793 |

Tab. 1: Accuracy of the course outcome using the previously defined test and validation set and the NB, kNN and SVM methods.

Note that the separation of the training, test and validation set was not replicated from [Ku15], since their weekly approach would have made an assessment of fairness overly complicated. The accuracy values are comparable to those of [Ku15]. All in all, the usage of the virtual learning environment had a high predictive value of the course outcome.

## 4 Results

All in all, we have implemented three machine learning algorithms for the prediction of the course outcome in real MOOC data and calculated their fairness measures: PP, EO and SA. In any of the three measures, a larger number means a higher predictive power and is thus advantageous. The bigger the difference between the subgroups for each of the measures, the more unfair could the model be considered. For example, in Table 2 we can see in the first row (predictive parity) that the difference between PPV values of students with and without disabilities is significant and that the PPV value of students with no disability (abbreviated as N) is higher.

| Fairness Definition | NB | | kNN | | SVM | |
|---|---|---|---|---|---|---|
| Predictive Parity (PP) | 0.032 (F) | **0.086** (N) | **0.084** (F) | **0.079** (N) | **0.085** (F) | **0.082** (N) |
| Equalised Odds (EO) | **0.372** | **0.210** | **0.074** | 0.022 | **0.064** | 0.006 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (M) | (N) | (M) | (D) | (M) | (N) |
| Slicing Analysis (SA) | **0.044** (M) | 0.020 (N) | 0.018 (F) | 0.002 (D) | **0.038** (F) | 0.002 (D) |

Tab. 2: Difference of PPV, TPV and AUC for the different prediction methods grouped by gender. In parenthesis, we denoted the group with the larger PPV, TPV, AUC values respectively by F/M for female/male and D/N for disability or no disability.

Most models show significant differences in accuracy between the sub-groups. Depending on the measure, model accuracy was sometimes higher for female and sometimes for male students. With one exception (EO for SVM), it was always better for students without disability. Interestingly, all models predicted below average pass-rates for female course participants, where they were higher in reality. NB and kNN models over-estimated pass-rate for disabled students, SVM strongly underestimated it. The reason for the differences in model fairness cannot be explained at this level of analysis, since the models work as black-box implementations.

## 5 Discussion

We have shown with the presented analysis on a real MOOC data set (OULAD) that, independently of the applied predictive model, predictive quality among learner sub-groups varies. The level of variation in difference of predictive quality between sub-groups was assessed with different fairness measures, with different, even contradictory results.

It should be noted that contradictory results mainly appear for the PP fairness measure, which might have to be calibrated differently (with a higher threshold for the 'unfair' category). Other than that, the EO measure quantified the NB method as highly discriminatory and the kNN and SVM methods only when grouping by gender. A reason for this could be the algorithms heavy reliance on relative frequencies in the original data.

When taking our exploratory analysis into account, the results of the SA method come closest to what we would expect, since no clear tendencies with respect to gender were visible. It should be noted that, when grouping for disability, the significant fairness measures consistently have higher PPV, TPV and AUC values for non-disabled students. This confirms our first expectation of possibly discriminatory data, which might lead to a "garbage in – garbage out" mechanism in the ML algorithms, especially using NB. Therefore, we can not fully conclude the validity of the fairness measures, but confirm that tendencies in the data were (at least in part) reproduced by the algorithms.

In order to get a more comprehensive view of the appropriateness of fairness measures in LA settings, further research and analyses on different datasets are required.

## Bibliography

[AP12] Arnold, K. E.; Pistilli, M. D.: Course signals at Purdue: Using learning analytics to increase student success: Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, 2012; S. 267–270.

[Ch17] Chouldechova, A.: Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. In Big data, 2017, 5; S. 153–163.

[FY15] Fei, M.; Yeung, D.-Y.: Temporal models for predicting student dropout in massive open online courses: 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE, 2015; S. 256–263.

[GBB19] Gardner, J.; Brooks, C.; Baker, R.: Evaluating the Fairness of Predictive Student Models Through Slicing Analysis, 2019.

[GD12] Greller, W.; Drachsler, H.: Translating learning into numbers: A generic framework for learning analytics, 2012.

[HPa16] Hardt, M.; Price, E.; and Srebro, N.: Equality of Opportunity in Supervised Learning. In (Lee, D. D. et al. Hrsg.): Advances in Neural Information Processing Systems 29. Curran Associates, Inc, 2016; S. 3315–3323.

[HZZ17] Hlosta, M.; Zdrahal, Z.; Zendulka, J.: Ouroboros: early identification of at-risk students without models based on legacy data: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. ACM, 2017; S. 6–15.

[KHZ17] Kuzilek, J.; Hlosta, M.; Zdrahal, Z.: Open university learning analytics dataset. In Scientific data, 2017, 4; S. 170171.

[Ku15] Kuzilek, J. et al.: OU Analyse: analysing at-risk students at The Open University. In Learning Analytics Review, 2015; S. 1–16.

[MD10] Macfadyen, L. P.; Dawson, S.: Mining LMS data to develop an “early warning system” for educators. In Computers & education, 2010, 54; S. 588–599.

[Mu12] Murphy, K. P.: Machine learning: a probabilistic perspective. MIT press, 2012.

[O’17] O’Neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books, 2017.

[Si13] Siemens, G.: Learning analytics: The emergence of a discipline. In American Behavioral Scientist, 2013, 57; S. 1380–1400.

[SP13] Slade, S.; Prinsloo, P.: Learning Analytics. In American Behavioral Scientist, 2013, 57; S. 1510–1529.

[VR18] Verma, S.; Rubin, J.: Fairness definitions explained: 2018 IEEE/ACM International Workshop on Software Fairness (FairWare). IEEE, 2018; S. 1–7.

.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 229

# Towards a Critical Design Agenda in Support of Collective Learning Ecologies

Christoph Richter and Heidrun Allert[1]

**Abstract:** This position paper seeks to chart a critical design agenda in support of collective learning ecologies, evolving assemblages of digital, spatial, social, cultural, and/or knowledge resources that are aimed to foster forms of collective learning and knowledge creation. Starting from an ecological perspective on learning, the paper challenges individualistic notions of learning and instrumental understandings of educational technology and argues that there is a need for educational formats and technologies capable to support learners engaged in epistemic endeavors that reach beyond established institutional boundaries and address them as responsible citizens. It is argued that there is a need (a) to take stock of new and alternative knowledge practices, (b) to reconsider the existing social, legal and technical protocols, standards, and infrastructures, (c) to cultivate social relations beyond institutional boundaries, as well as (d) to work towards a pedagogy of articulation and risk.

**Keywords:** critical design, learning ecologies, educational technologies, computer-supported collaborative learning, inquiry-based learning, epistemic practices

## 1 An Ecological Perspective on Learning

Recently there has been an increased interest in what might be called an ecological perspective on learning, e.g. [Ja13]. Broadly speaking an ecological perspective on learning draws attention to the evolving assemblages of digital, spatial, social, cultural, and/or knowledge resources learners are creating and are making use of in their study work. In a narrower sense, an ecological perspective challenges individualistic notions of learning and education but "conceives of learning as an irreducible, mutually constitutive set of relationships between individuals and their social and material environments" [DJ17]. Besides conceptual contributions, aimed to substantiate the notion of learning ecologies theoretically, e.g. [Ke12], the interest in an ecological take on learning is also mirrored in pedagogical and technical efforts aimed to device Personal Learning Environment, e.g. [Wi07], to support processes of community-based learning, e.g. [FGW07], and/or to foster processes of mass collaboration, e.g. [Je17]. In all these cases pedagogical and technological means are supposed to transcend institutional, disciplinary, social, cultural, or technical boundaries and allow for new forms of learning and epistemic engagement. While an ecological perspective on learning has opened up new strands of research and development for educational technologies such as the advancement of Massive Open Online Courses (MOOCs) or the configuration of tailored learning environments, the pedagogical and technical design issues associated with these

[1] Christian-Albrechts-Universität zu Kiel, Abteilung Medienpädagogik / Bildungsinformatik, Olshausenstr. 75, 24116 Kiel, {richter,allert}@paedagogik.uni-kiel.de

doi:10.18420/delfi2019_307

developments are frequently approached from an instrumental perspective only. The availability of new technologies, networked knowledge artifacts, and human resources are primarily seen as means to improve learning outcomes and to ensure that students become proficient knowledge workers skilled for the digital age. However, such an instrumental perspective does neither account for the cultural dimension of education and knowledge work, nor does it advance a critical understanding of the challenges, learners are facing when trying to develop their own learning ecologies. Against this background we argue, that to leverage an ecological perspective on learning and to become serious about learners as agentic actors who pursue their own epistemic projects, we need to take a critical perspective on our design efforts,cf. [RA17]. This holds for the field of educational technology in general as well as the narrower domain of computer-supported collaborative learning. We also have to reconsider our role as educators, technologists, and researchers, in order to advance educational formats and technologies capable to support learners engaged in epistemic endeavors that reach beyond institutional boundaries and address them as responsible citizens. Our intent hence is to broaden the current discourse on educational and technological design. To illustrate the timeliness of such an effort and to outline a critical design agenda, we will draw on two ongoing R&D projects targeted at new collaborative forms of epistemic engagement. While the first project[2] is based on a participatory research agenda aimed to engage citizens of different age groups in an autoethnographic inquiry into their own cultural and aesthetic practices, the second project[3] is aimed to foster crowd-based research activities on sustainability among students in higher education. Despite the different subject domains, educational formats, and target populations, both projects can be understood as attempts to enrich the participants' learning ecologies, by introducing new technical, social, cultural, and epistemic resources that transcend institutional, disciplinary, and physical boundaries.

## 2 Understanding Digital Technologies – Types of Mediation

To advance an understanding of the complexities entailed in an ecological perspective on learning, the critical design agenda we are proposing builds on the multidimensional model of mediation suggested by [BR00] and its later adaptation to the field of computer-supported collaborative learning by [PEH12]. In a nutshell the model holds that (digital) technologies comprise of a material and/or symbolic artifact, as well as associated socio-cultural schemes on how to understand and make use of these artifacts as part of a situated practice. As a consequence, and in contrast to an instrumental conception of technology, (digital) technologies are not simply means to towards a given end, but inescapably mediate practice along multiple dimensions simultaneously. As suggested by [PEH12], educational technologies can be described and analyzed along at least four dimensions or types of mediation. These types of mediation include (a) epistemic mediation, the situated procedures through which knowledge is created collaboratively, (b) pragmatic mediation,

[2] Onlinelabor für Digitale Kulturelle Bildung (http://www.digitalekultur.online)
[3] SCoRe – Student Crowd Research (https://scoreforschung.com)

the situated procedures through which the collective efforts are coordinated and regulated, (c) social (or collaborative) mediation, the situated procedures through which social relations and networks are (reproduced), as well as (d) reflective mediation, the situated procedures through which the collective efforts are assessed, evaluated, and advanced. For example, a collaborative word editor not just transforms the way ideas are collaboratively articulated, it also alters the way respective processes are coordinated, how social relations are formed, how measures of accountability and quality are re-produced. Bearing in mind that the 'multimediation' of (educational) practice is intrinsic to any kind of technology, we will use the dimension of mediation suggested by [PEH12] in the following as a framework to organize the critical design agenda we are suggesting.

## 3 Taking Stock of New and Alternative Knowledge Practices

The first issue relates to the epistemic dimension of the learning process and refers to the rules, conventions, and normative commitments that are constitutive for the contexts we are acting in and designing for. It refers to the "epistemic frames" [Sh06] enacted in a particular setting as well as the „ways of knowing that are taken to be legitimate, consequential, worthy of discussion, and useful for justifying actions by people engaged in accomplishing some concerted task" [Su02]. While institutions in (higher) education are still very much driven by disciplinary bodies of knowledge, as well as specific knowledge practices and methodologies, these might be of little or no relevance if we leave the realm of academic institutions, whether applied or not. In our current projects we are actually facing a multitude of new and alternative knowledge practices, including for example new forms of crowd-based investigation, activist and collective artistic research, as well as new tribes of (auto-)biographic and (auto-)ethnographic inquiry. If our intent is to support learning ecologies that transcend institutional and disciplinary boundaries, we hence have to develop new sensitivities to the polyphony of epistemic practices not only within but also outside of academia and find ways to accommodate for different epistemic forms and games. This not only calls for new educational formats, but also asks us to reconsider the ways in which epistemic practices are implicated in digital and networked technologies. Towards this end, it is crucial (a) to reflect on and articulate not only the explicit, but also the implicit epistemic scaffolds and models that are built into respective technologies, as well as (b) to inquire into the modes of knowledge and articulation that are rendered perceptible or imperceptible by a digital technology.

## 4 Reconsidering the Protocols, Standards, and Infrastructures we Build upon

The second issue is concerned with the pragmatic mediation of the learning processes, the organization, coordination, and orchestration of the epistemic efforts. It entails the social and technical mechanisms that grant or restrict access to technical, social, cultural, and

epistemic resources, the means by which these resources can be created, mobilized, appropriated, or repurposed, but also the setup and characteristics of the institutional and technical infrastructures that are made use of. While a focus on personal learning ecologies implies a shift from institutionalized to more decentralized modes of organization, coordination, and orchestration, it requires us to reconsider the ways in which students' learning ecologies are implicated in social and technical systems and therefore subject to respective forms governance and politics, cf. [Kn13]. For example, aiming to provide students and citizens with a technological infrastructure that allows them to pursue their own (collective) epistemic interests, brings immediately forward questions of authorship, intellectual property rights, the interoperability with other systems, the sustained availability of the resources and technologies provided, as well as the extent to which the resources can be reused and repurposed, not only legally but also technically. Hence, there is a need to think not only about particular resources and how these can be integrated into a learning ecology, but also the social, legal, and technical protocols, standards, and infrastructures, in which these ecologies take shape, cf. [Ma16]. In particular, it is important (a) to devise social and technical protocols that transcend institutional boundaries as well as the confines of commercial vendors, (b) to envision systems that are capable to adapt to the evolving coordinative needs in open ended collaborative efforts, and (c) to balance individual accountability with new forms of collective authorship.

## 5 Cultivating Social Relations beyond Institutional Boundaries

The third and closely related issue touches upon the social mediation, the seeding and cultivation of social relations and networks in support of the learner's epistemic efforts. Locating learning not within the boundaries and formats of educational institutions but striving for more inclusive, boundary-crossing, and networked forms of learning does not simply imply access to a broader and more heterogeneous set of (social) resources, but also entails a shift in power-relations. Yet, such a shift is not a unidirectional effort of empowerment, but a highly dynamic and reflexive process. This holds for both, the relations between the learners and the educational institutions, as well as the relations between those engaged in the learning ecologies. For example, asking citizens of different age groups to take part in a research project as knowledgeable experts of the cultural and aesthetic practices they are engaged in challenges established role-models, social expectations, responsibilities, and obligations. Similarly asking students to actively contribute to research on sustainability raises questions on how these efforts will interfere with established or emerging forms of scientific and public deliberation and discourse. Accordingly, there is a need (a) to think of social and technical arrangements that are open to all those who have an interest in what is at stake, but also (b) to devise for new forms of participation and collaborative deliberation on issues that matter to the actors involved.

## 6 Towards a Pedagogy of Articulation and Risk

The fourth issue, finally, relates to the reflective mediation of epistemic endeavors, the processes through which respective practices are articulated, reflected, and advanced. It refers to the question of the (educational) outcomes that are deemed desirable and worthwhile, and the ways in which respective aims are negotiated and established. While from an institutional perspective the desirability of educational outcomes is usually determined by established curricula and educational policies, these are not necessarily relevant to or might even be in conflict with the aspirations of a learner and the setup of his or her learning ecology. For example, citizens' inquiries into their own cultural and aesthetic practices, or students' research into sustainability issues might less be oriented towards the achievement of preestablished learning goals than driven by their curiosity or ambition to solve a particular problem. If we are serious about learners as deliberate actors and citizens, desired outcomes cannot be pre-defined by educational institutions alone, but that there is a need for mechanisms that allow the participants to articulate and negotiate their (common) concerns and the kind of outcomes they deem worthwhile and desirable. As suggested by [Gr97], such efforts might call for a new model of pedagogy, "a pedagogy of articulation and risk", that refuses "to assume ahead of time that it knows the appropriate knowledge, language or skills, […] a contextual practice that is willing to take the risk of making connections, drawing lines, mapping articulations, between different domains, discourses, and practices, to see what will work, both theoretically and politically". It also challenges instrumental notions of technology as it requires to have a closer look into the processes through which technologies emerge and the politics they entail.

## 7 Outlook

While our concern for these four issues is clearly grounded in our current project efforts, we believe that they might form a starting point for a critical design agenda in the fields of educational technology as well as CSCL. As far as we can see, related issues have popped up recurrently in various places, but there has been no systematic debate on the wider pedagogical, societal, political, and technical implications of our efforts. Therefore, we understand this contribution as an attempt to broaden the discourse, but also see it as an opportunity to enter into new strands of design, research, and technological development aimed to foster the advancement of more sustainable learning ecologies.

**Acknowledgments.** This research has received partial funding from the Federal Ministry of Education and Research under grant agreement n° 01JKD1705 and 16DHB2119.

Bibliography

[BR00] Béguin, P.; Rabardel, P.: Designing for instrument-mediated activity. Scandinavian Journal of Information Systems, 12, pp. 173-190, 2000.

[DJ17] Damşa, C.I.; Jornet, A.: Revisiting Learning in Higher education—Framing notions redefined through an Ecological Perspective. Frontline Learning Research Journal, 4/2, pp. 12– 20, 2018.

[FGW07] Fischer, G.; Rohde, M.; Wulf, V.: Community-based learning: The core competency of residential, research-based universities. International Journal of Computer-Supported Collaborative Learning, 2/1, S. 9–40, 2007.

[Gr97] Grossberg, L.: Bringing It All Back Home. Pedagogy and Cultural Studies. In (Grossberg, L., ed.): Bringing It All Back Home. Essays on Cultural Studies. Duke University Press, Durham, pp. 374–390, 1997.

[Ja13] Jackson, N J.: The concept of learning ecologies. In (Jackson, N.J.; & G. B. Cooper, G.B. eds.): *Lifewide learning, education and personal development e-book*, pp. 1–21, 2013, retrieved from http://www.lifewideebook.co.uk/uploads/1/0/8/4/10842717/chapter_a5.pdf, 22/06/2019.

[Je17] Jeong, H. et. al.: Joint interactions in large online knowledge communities: The A3C framework. International Journal of Computer-Supported Collaborative Learning, 12/2, pp. 133–151, 2017.

[Ke12] Kemmis, S. et al.: Ecologies of Practices. In (Hager, P.; Lee, A.; Reich, A. eds.): Practice, Learning and Change - Practice-Theory Perspectives on Professional Learning. Springer, Dordrecht, pp. 33–49, 2012.

[Kn13] Knox, J.: The limitations of access alone: Moving towards open processes in education technology. Open Praxis, 5/1, pp. 21-29, 2013.

[Ma16] Mansour, E. et al.: A demonstration of the solid platform for social web applications. In Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada 2016. ACM press, New York, pp. 223-226, 2016.

[PEH12] Paavola, S.; Engeström, R.; Hakkarainen, K.: The Trialogical Approach as a New Form of Mediation. In (Moen, A.; Mørch, A.I.; Paavola, S. eds.): Collaborative Knowledge Creation: Practices, Tools, Concepts. Sense Publishers, Rotterdam, pp. 1-14, 2012.

[RA17] Richter, C.; Allert, H.: (2017). Design as Critical Engagement In and For Education. EDeR - Educational Design Research, 1/1, pp. 1-20, 2017.

[Sh06] Shaffer, D.W.: Epistemic frames for epistemic games. Computers & Education, 46, pp. 223-234, 2006.

[Su02] Suchman, L.A.: Practice-Based Design of Information Systems: Notes from the Hyperdeveloped World. The Information Society: An International Journal, 18/2, pp. 139-144, 2002.

[Wi07] Wilson, S. et al.: Personal Learning Environments: Challenging the dominant design of educational systems. Journal of E-Learning and Knowledge Society, 3/2, pp. 27–38, 2007.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 235

# PLAY & CODE – Eine Serious Game basierte Lernplattform

Elias John[1]

**Abstract:** Viele Positionen im Informatiksektor können schon heute nicht besetzt werden und es ist davon auszugehen, dass dieser Mangel an Informatiker*innen in Zukunft noch steigen wird. Problematisch ist dabei auch die hohe Abbrecherquote in den entsprechenden Studiengängen. Viele Themen des Studiums erscheinen als zu theoretisch, abstrakt oder nicht praxisrelevant. Es stellt sich daher die Frage, wie Lerninhalte adäquater vermittelt und mehr Schüler*innen für Themen der Informatik begeistert werden können. Als Mittel zur Steigerung der Motivation scheinen besonders Spiele geeignet. Der vorliegende Aufsatz soll daher folgende Frage beantworten: Können mithilfe eines Serious Games basierten Lernsystems Motivation und Lernerfolg für grundlegende Programmierkenntnisse gesteigert werden?

**Keywords:** Serious Games, Grundlagen der Programmierung, JavaScript, Lernplattform

## 1 Einleitung

Durch die fortschreitende Digitalisierung besteht ein hoher Bedarf an Informatiker*innen. Trotz einer zunehmenden Anzahl an Informatik-Studierenden steigt die Zahl der unbesetzten Stellen seit 2014 pro Jahr stetig an [IW17]. Dieser Mangel an Fachkräften wird zudem durch eine hohe Abbrecherquote im Studium verstärkt [He17]. Nach Muratet et al. [Mu09] empfinden viele Studierende zu Beginn des Studiums das Themengebiet als zu theoretisch und abstrakt. Wie können also Lerninhalte praxisnäher vermittelt sowie bereits Schüler*innen für Themen der Informatik begeistert werden? Als Mittel für die Steigerung der Motivation scheinen besonders Spiele geeignet, da sie einen leichten und niedrigschwelligen Zugang zum jeweiligen Thema ermöglichen. Es stellt sich damit die Frage, ob mithilfe eines Serious Games basierten Lernsystems Motivation und Lernerfolg für grundlegende Programmierkenntnisse gesteigert werden können.

Dafür wird im folgenden Kapitel 2 zunächst auf den Stand der Forschung eingegangen. Bestehende Serious Games und ihre Eigenschaften werden vorgestellt. Anschließend wird im dritten Kapitel näher auf das Konzept der Serious Game basierten Lernplattform PLAY & CODE eingegangen. Diese vermittelt grundlegende Programmierkonstrukte mittels Serious Games und passt die Texte individuell an den/die Benutzer*in an. Um die Forschungsfrage zu beantworten, wurde die Plattform an mehreren Schulen evaluiert. Die

[1] Humboldt-Universität zu Berlin, Institut für Informatik, Rudower Chaussee 25, 12489 Berlin, elias.john@student.hu-berlin.de

doi:10.18420/delfi2019_248

Methodik und Ergebnisse der Evaluation werden im vierten Kapitel vorgestellt. Abschließend wird die Forschungsfrage beantwortet und ein Fazit gezogen.

## 2 Stand der Forschung

Für ein erfolgreiches Lernen ist die intrinsische Motivation des/der Lernenden eine wichtige Voraussetzung [Hu07]. Als ein erfolgsversprechendes Mittel zur Steigerung der Motivation gelten Spiele [Dö16]. Solche Serious Games werden in verschiedenen Bereichen, wie beispielsweise der Biologie und Chemie, eingesetzt (vgl. [Ch15]). Exemplarisch werden einige solcher Serious Games aus dem Bereich der Informatik vorgestellt. Dabei lässt sich feststellen, dass die Programmierung auf verschiedene Weise (textuell oder grafisch) vermittelt und in verschiedenen Umgebungen eingebettet werden kann (unterrichtsbegleitend, alleinstehend, sowie als eigenständiges Spiel oder eingebettet in einer Lernumgebung).

Das Spiel **„Program your Robot“** von Kazimoglu et al. [Ka12] ist ein Serious Game zum Üben grundlegender Programmierkonstrukte, wie konditionaler Logik und Schleifen. Den Lernenden wird dabei vor allem vermittelt, wie Algorithmen aufgebaut und Programme systematisch auf Fehler untersucht werden können. Ziel des Spiels ist es, einem Roboter bei der Flucht zu helfen. Dafür muss der/die Spieler*in einen Lösungsalgorithmus entwickeln, indem er vorgefertigte grafische Befehlssymbole in beliebiger Reihenfolge in eine begrenzte Anzahl von Slots zieht. Zufällig erzeugte Belohnungsitems steigern den Punktestand der Spieler*in.

**„Prog&Play“** ist ein Serious Game von Muratet et al. [Mu09] zur Übung von Programmierung. Das Serious Game ist begleitend zum Unterricht durch eine Lehrkraft gedacht und enthält keine Lehrinhalte. Das Spielgeschehen findet im Inneren eines gehackten Computersystems statt, über welches der/die Spieler*in in mehreren Leveln mit dem Steuern von Elementen wie Bits und Bytes wieder die Kontrolle erlangen muss. Die Lösung für das jeweilige Problem wird mittels einer zur Verfügung gestellten API programmiert, die durch eine Vielzahl an Programmiersprachen angesteuert werden kann.

Schäfer et al. [Sc13] stellen eine **kollaborative, Multitouch Lernumgebung** vor. Ziel ist es, gemeinsam an einem Multitouch-Gerät mathematische Logik zu erlernen und zu üben. Nach den Autor*innen führt Kollaboration zu mehr Interaktion und einer besseren Aufnahme des Lernstoffes. Neue Lerninhalte werden zunächst grafisch animiert erklärt. Anschließend werden Aufgaben von dem/der Benutzer*in selbst ausgeführt. Dabei gibt das System Feedback und stellt Quizfragen zum Thema. Im Freeplay-Abschnitt können weitere Aufgaben gelöst werden. Hierbei wird der/die Benutzer*in mit Punkten für schnelle Reaktionen belohnt, welche für andere Spielende einsehbar sind. Dadurch sollen sich andere Spieler*innen motiviert fühlen, bessere Werte zu erreichen.

Von den genannten Arbeiten gehen nur Schäfer et al. [Sc13] näher auf die verwendeten didaktischen Modelle und Theorien ein. Dies scheint ein generelles Phänomen zu sein: In

53 untersuchten Veröffentlichungen zu Serious Games von 2002 bis 2013 stellten Cheng et al. [Ch15] fest, dass nur die Hälfte der Autor*innen den Einsatz solcher Spiele lerntheoretisch begründen. Zudem wurde die Effizienz von Serious Games empirisch noch nicht vollständig nachgewiesen: So stellen beispielsweise Girard et al. [GEM13] in ihrer Meta-Analyse zu Serious Games diverse methodische Mängel bei der Untersuchung der betrachteten Artikel fest. Serious Games scheinen ein hilfreiches Werkzeug für Lernzwecke zu sein, jedoch haben bisherige Evaluationsstudien noch keine hinreichende Begründung für den Erfolg des Einsatzes von Serious Games geliefert. Im Folgenden wird daher eine didaktisch fundierte Serious Game basierte Plattform zum Erlernen von Programmierung vorgestellt.

# 3 Konzept der Lernplattform PLAY & CODE

## 3.1 Didaktik der Lernplattform

Didaktisch gründet die im folgenden vorgestellte Lernplattform auf dem ARCS-Modell von Keller [Ke87] sowie den „First Principles of Instruction“ von Merrill [Me02]. Das Modell von Merrill basiert auf einer umfangreichen Untersuchung verschiedener Instruktionsmodelle und versucht die Ansätze in einem Modell zu vereinen. Die Arbeit Merrills fußt auf dem problembasierten Lernansatz, welcher sich als sehr effektiv erwiesen hat (vgl. [Dö16] und [Sa06]). Durch den Problembezug wird der Lerninhalt konkretisiert und wirkt so weniger abstrakt, was sich unter anderem positiv auf die Motivation der Lernenden auswirken kann. Zusätzlich wird das ARCS-Modell von Keller berücksichtigt, welches verschiedene Aspekte zur Motivationserzeugung beschreibt. Der Einsatz eines Serious Game als Lehrmittel trägt zum Spaß beim Lernen bei und ruft intrinsische Motivation hervor. Da Programmieraufgaben auf verschiedene Wege gelöst werden können, wird außerdem Kreativität beim Lernen befördert.

Im Gegensatz zu anderen spielebasierten Lernplattformen, wie „Program your Robot“ von Kazimoglu et al. [Ka12], wird hier keine grafische, sondern die textuelle Programmiersprache JavaScript eingesetzt. Grafische Programmiersprachen scheinen einen sehr leichten Zugang zur Programmierung zu erlauben. Im Hinblick auf die von Merrill [Me02] und Keller [Ke87] geforderte Realitätsnähe und Alltagstauglichkeit wurde bei PLAY & CODE jedoch nach einer Sprache gesucht, welche im professionellen Bereich eine starke Verbreitung findet. Hierfür scheint sich JavaScript besonders zu eignen. So listet beispielsweise der TIOBE Programming Community Index [TI19] JavaScript unter den Top zehn Programmiersprachen. Die Schüler*innen können ihr erlerntes JavaScript-Wissen nach dem Lernkurs direkt und vielseitig in Web-, Desktop- und Serveranwendungen verwenden.

## 3.2 Struktur des Systems

Die Lernplattform ist als ein Online-Mehrbenutzersystem konzipiert, bei der sich Interessierte selbstständig anmelden können.[2] Nach dem erstmaligen Login muss der/die Benutzer*in zunächst einen kurzen Fragebogen zu Programmier-Vorerfahrungen beantworten. Dabei wird mit Likert-Skalen von 1 (keine Erfahrung) bis 5 (sehr erfahren) nach Programmiererfahrungen im Allgemeinen und Erfahrungen mit JavaScript im Speziellen gefragt. Diese Informationen werden zur Errechnung eines Erfahrungswertes verwendet, welcher die Länge der Lehrtexte beeinflusst. Hierauf wird später detaillierter eingegangen.

Der/Die Benutzer*in kann aus einer Kursübersicht seine/ihre Kurse frei auswählen, welche mit einem kurzen Text und einem Thumbnail beschrieben sind. Hier kann ein Kurs entweder begonnen oder zu einem späteren Zeitpunkt auch fortgesetzt werden. Jeder Kurs ist in mehrere Lektionen gegliedert, welche der Reihe nach abgeschlossen werden müssen.

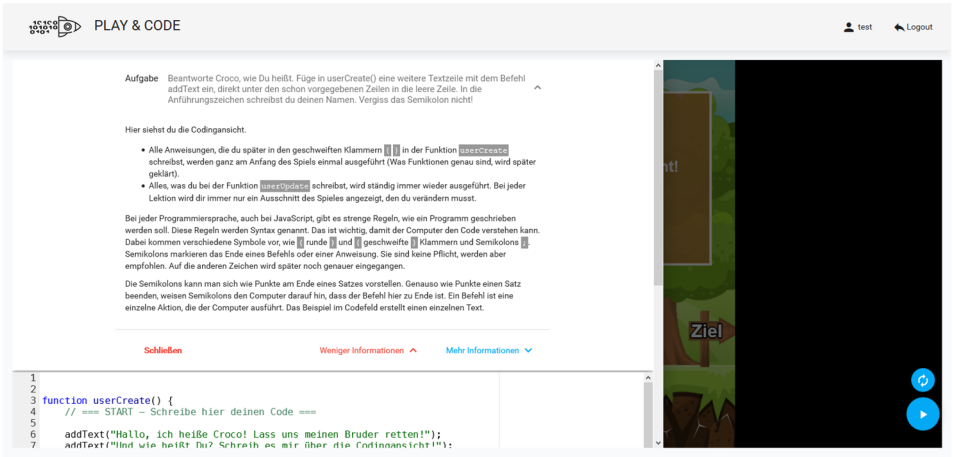



Abb. 1: Screenshot aus dem Kurs "Crocos Bruder" mit geöffneter Coding-Ansicht.

Jeder Kurs stellt ein Videospiel dar, mit welchem eine in sich abgeschlossene Geschichte erzählt wird. Die Lektionen des Kurses sind die einzelnen, aufeinanderfolgenden Abschnitte dieses Spiels. In jeder Lektion wird in einer animierten Intro-Sequenz die Geschichte weitererzählt und der Protagonist vor ein neues Problem gestellt. Hiermit soll sich der/die Benutzer*in in das Spielgeschehen integriert fühlen. Auf die Intro-Sequenz folgt die Lern- und Programmierphase, in welcher der/die Benutzer*in neue Inhalte erlernt. Dafür muss zunächst die Coding-Ansicht geöffnet werden, in welcher die Lehrtexte, die Aufgabe und die Codeeingabe angezeigt werden. Die Texte können auch Bilder und formatierte Texte enthalten (vgl. Abb. 1). Der Text unterteilt sich in folgende Absätze:

[2] Implementiert wurde die Lernplattform PLAY & CODE mit dem Java-Framework Spring Boot für das Backend und mit dem Frontend-Framework Angular 6. Das Backend verwaltet dabei alle relevanten Daten in einer MySQL-Datenbank und stellt diese über eine REST-API für das Frontend bereit. Dabei werden Abfragen authentifiziert und autorisiert.

1.: Grundlegende Informationen zum jeweiligen Lernthema werden bereitgestellt. 2.: Informationen werden weiter ausgeführt und Konstrukte detaillierter erklärt. 3.: Konkrete Positiv- und Negativbeispiele werden aufgelistet und weiterführende Ressourcen verlinkt.

### 3.3 Adaptivität der Lehrtexte

Da jeder/jede Lernende einen eigenen Lernstil und unterschiedliche Voraussetzungen mitbringt, ist es notwendig, auf diese individuell einzugehen. Diverse Studien haben nach Vandewaetere et al. [VDC11] gezeigt, dass dieser individuumsbasierte Ansatz uniformen Modellen, die die Bedürfnisse der Lernenden kaum berücksichtigen, überlegen ist. Der Lehrtext wird daher an den/die Benutzer*in adaptiert: Je nachdem, wie hoch der errechnete Erfahrungswert der Benutzer*innen ist, wird ein unterschiedlich großer Ausschnitt des gesamten Textes angezeigt. Der jeweilige Erfahrungswert der Benutzer*innen ist auf einer Skala von 1 bis 15 angesiedelt. Bei einem Wert unter 6 werden alle drei Abschnitte angezeigt: Der/die Benutzer*in benötigt detaillierte Informationen und konkrete Beispiele, um die Aufgabe lösen zu können. Insbesondere Positiv- und Negativbeispiele sind nach Merrill [Me02] hilfreich, Inhalte zu veranschaulichen. Bei einem Erfahrungswert zwischen 6 bis 10 werden zwei Abschnitte angezeigt. Bei Werten darüber wird lediglich der erste Abschnitt dargestellt: Der/Die Benutzer*in kennt bereits die Paradigmen aus anderen Sprachen und benötigt lediglich knappe Informationen über die JavaScript spezifische Verwendung. Meint der/die Benutzer*in mehr oder weniger Informationen zu benötigen, kann er/sie den Text weiter ein- bzw. ausklappen. Der Erfahrungswert errechnet sich aus den Informationen des Fragebogens, den benötigten Versuchen in der aktuellen Lektion und den manuellen Veränderungen des Textausschnitts. Der Wert dient dabei lediglich zur Berechnung und wird dem/der Benutzer*in nicht angezeigt. Mit diesem mikroadaptiven Ansatz[3] (vgl. [VDC11]) soll Benutzer*innen mit Vorerfahrungen zu viel Text erspart bleiben und Anfänger*innen die benötigten Informationen zur Verfügung gestellt werden.

### 3.4 Spielablauf

Nachdem Lehrtexte und Aufgabenstellung von dem/der Benutzer*in gelesen wurden, kann er/sie mit dem neu gewonnenen Wissen eine Lösung für das Problem programmieren. Der für die Spielmechanik benötigte Code wird dabei in – für den/die Benutzer*in nicht sichtbare – Dateien ausgelagert. Die Benutzer*innen sehen lediglich die für sie bestimmte Datei UserCode, welche befüllt werden muss, um die Aufgabe zu lösen. Damit soll sichergestellt werden, dass der/die Benutzer*in nicht von der Codefülle überfordert ist, sondern sich auf das Wesentliche konzentrieren kann. Die Datei UserCode enthält initial zwei Funktionen: userCreate und userUpdate. userCreate wird einmalig bei der

[3] Im Gegensatz zu makroadaptiven Ansätzen, bei denen der/die Lernende einmalig klassifiziert und Inhalte darauf basierend bereitgestellt werden, werden die Inhalte beim mikroadaptiven Ansatz dynamisch während der Kursdurchführung angepasst [VDC11]. Bei Letztem werden also individuelle, möglicherweise temporäre Eigenschaften des/der Lernenden stärker berücksichtigt und damit auf Mikroebene agiert.

Initialisierung des Spiellevels ausgeführt, während userUpdate fortlaufend in der Spielschleife einmal pro Bilddarstellung (frame) aufgerufen wird. Die Funktionen können leer sein oder vorgegebene Codes enthalten, die von dem/der Benutzer*in ergänzt werden müssen.

Nachdem der/die Benutzer*in seinen/ihren Lösungscode geschrieben hat, kann er/sie auf „Play“ drücken. Sollten syntaktische Fehler, wie zum Beispiel falsch gesetzte Klammern, vorhanden sein, wird die Ausführung unterbrochen und die Fehler werden mit Zeilennummern und Fehlermeldung angezeigt. Andernfalls wird das modifizierte Level gestartet. Dies entspricht der Empfehlung Merrills [Me02], dass der/die Benutzer*in erworbenes Wissen unmittelbar anwenden kann und Feedback erhält. Der/Die Benutzer*in kann dabei das Problem auf verschiedene, kreative Weise lösen, solange die semantischen Anforderungen erfüllt sind. Diese werden vom Kursautor oder von der Kursautorin definiert. Einerseits können Werte und Zustände, wie die Position der Spielfigur, überprüft werden. Andererseits kann auch die Codeabgabe auf bestimmte Schlagwörter untersucht werden.

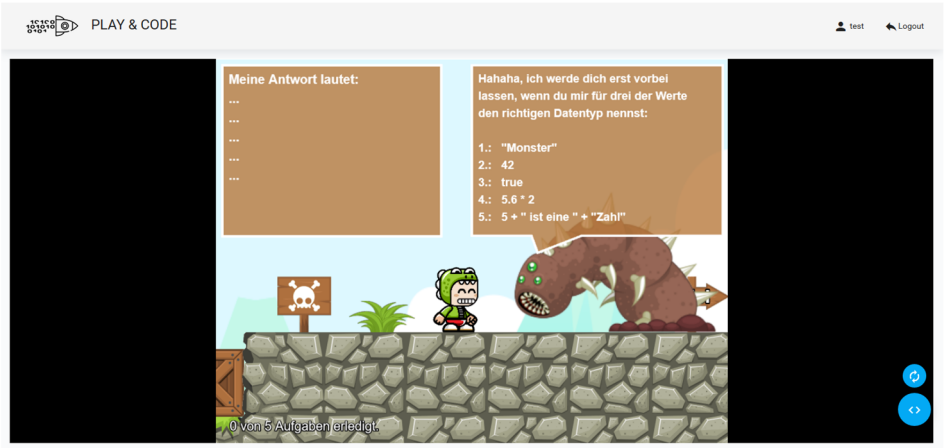



Abb. 2: Screenshot aus dem dritten Level von "Crocos Bruder"

Durch das Lösen weiterer Probleme durch zusätzliche, optionale Programmierung erscheinen in einigen Leveln Diamanten, welche der/die Benutzer*in einsammeln kann. Je nachdem ob und wie viele Diamanten eingesammelt wurden, wird dem/der Benutzer*in nach Levelabschluss eine Bronze-, Silber oder Goldmedaille vergeben. Dadurch soll der Erfolg sichtbar gemacht und eine Zufriedenheit der Benutzer*innen erreicht werden (vgl. Kellers Satisfaction [Ke87]). Im unteren Bildschirmbereich wird ihm/ihr die Gesamtanzahl der zu lösenden Teilprobleme und der aktuelle Fortschritt angezeigt (siehe Abb. 2). In Leveln ohne Diamanten werden Goldmedaillen vergeben. Im Benutzerprofil können gewonnene Trophäen betrachtet werden. Zusätzlich wird ein Prozentsatz von Spieler*innen angezeigt, die diese Trophäe ebenfalls erhalten haben. Diese Gamification-Elemente sollen zum spielerischen Wettstreit führen und dadurch die Motivation zusätzlich steigern.

Zur Demonstration und Evaluierung der Lernplattform wurde der Beispielkurs „Crocos Bruder“ implementiert, mit welchem grundlegende Programmierkonzepte von JavaScript für Programmier- und JavaScript-Anfänger*innen vermittelt werden. Der/Die Benutzer*in spielt dabei ein 2D-Adventure-Spiel und muss dem Protagonisten Croco bei der Rettung seines Bruders helfen, welcher zu Beginn der Geschichte entführt wird und im letzten Level des Spiels gerettet werden kann. In jedem Level wird Croco vor ein neues Problem gestellt, welches auf den vorangegangenen Leveln aufbaut und mit den erlernten Programmierkonstrukten lösbar ist. Das neu erworbene Wissen baut so auf dem vormals erlangtem auf (vgl. [Me02]) und die Relevanz dessen wird durch konkrete Anwendung und Kontextbeziehungen klar (vgl. Kellers Relevance [Ke87]). Folgende JavaScript-Themen werden in „Crocos Bruder“ behandelt: Syntax und Struktur, Kommentare, Datentypen (strings, numbers, booleans), Variablen, If und If-Else, Komparatoren, logische Operatoren, Funktionen, for-Schleifen. Die Themenwahl, deren Reihenfolge und die Lehrtexte sind an das Lehrbuch „JavaScript for Kids: A Playful Introduction to Programming“ [Mo14] angelehnt und für die Lernplattform angepasst. Die Themen decken zudem einen Teil der von Kazimoglu et al. [Ka12] und Muratet et al. [Mu09] identifizierten Programmiergrundlagen, wie konditionale Logik und Schleifen, ab.

## 4 Methodik und Ergebnisse der Evaluation

Um die eingangs vorgestellte Forschungsfrage zu beantworten, wurde das Lernsystem PLAY & CODE an mehreren Schulen in Berlin mit Schüler*innen unterschiedlichen Alters und mit wenig JavaScript-Erfahrungen evaluiert. Quantitative Daten wurden schriftlich durch standardisierte Online-Fragebögen und zudem automatisiert durch das System erfasst. Um den Lernerfolg der Schüler*innen feststellen zu können, wurde ein Pre-/Posttest durchgeführt. Hierbei wurden Fragen (16 Fragen im Pretest, 17 Fragen im Posttest) zu den grundlegenden Themenbereichen gestellt, welche auch im evaluierten PLAY & CODE Kurs „Crocos Bruder“ behandelt wurden (vgl. Kapitel 3.5). Im Pretest wurde zudem nach dem Alter, dem Geschlecht und der Spielhäufigkeit pro Woche bzw. Monat gefragt. Nach Abschluss des Posttests füllten die Schüler*innen einen Evaluationsfragebogen aus. Während der Benutzung von PLAY & CODE wurde außerdem die Bearbeitungsdauer und die Versuchsanzahl pro Lektion pro Benutzer*in gespeichert. Die Fragebögen (Fragen und Antwortskalen) wurden selbst erstellt, da keine getesteten Skalen hierzu vorliegen.

Im Zeitraum vom 15.02. bis zum 09.03.2018 haben sich bei PLAY & CODE 140 eindeutige Benutzer*innen registriert.[4] Die Zusammenführung von Pretest, Posttest und den Evaluationsfragebogen ergab 50 eindeutige, bereinigte Fälle. Hierbei handelt es sich um Benutzer*innen im Alter von 12 bis 18 Jahre. Um die Differenz zwischen Pre- und Posttest festzustellen, wurde zu den Antworten jedes Themengebiets in beiden Datensätzen ein gerundeter Durchschnittswert ermittelt. Hatte der/die Schüler*in die Hälfte oder mehr der

[4] Diese Zahl enthält auch Benutzer*innen, die nicht an der Befragung teilgenommen hatten und unaufgefordert oder auf Empfehlung von Anderen die Seite besuchten.

Antworten richtig beantwortet, wurde das Themengebiet als korrekt bearbeitet markiert, andernfalls als falsch. Aus den so entstandenen acht Variablen pro Datensatz wurden Summen berechnet, die über den Anteil von „richtig“ beantworteten Themenbereichen Aufschluss geben. Die Differenz stellt den Wissenszuwachs dar.

Im Evaluationsfragebogen wurden außerdem Fragen zur Usability gestellt. Die Bedienbarkeit wurde dabei mit durchschnittlich 1,9 als „gut“ bewertet. Das Design wurde von den Benutzer*innen im Mittel mit 1,4 bewertet. Insgesamt wurde PLAY & CODE mit einer 1,7 bewertet. Danach gefragt, ob die Schüler*innen gerne weitere Programmierinhalte mit Spielen lernen würden wollen, gaben 92 Prozent (46 Teilnehmende) „Ja“ an, 8 Prozent gaben „Nein“ an.

### 4.1 Qualität und Adaptivität der Lehrtexte

Die Lehrtexte sind bei PLAY & CODE essenziell für den Lernerfolg der Benutzer*innen. Es stellte sich daher die Frage, wie die Qualität die Lehrtexte des Evaluationskurses „Crocos Bruder“ eingeschätzt wird und wie sich diese auf den Lernerfolg der Benutzer*innen auswirkt. PLAY & CODE passt zudem die Länge der Lehrtexte an den geschätzten Wissensstand der Benutzer*innen an. Es sollte daher überprüft werden, wie die Textlänge bewertet wurde. Die Datenanalyse ergab, dass die Textlänge von 72 Prozent der Schüler*innen als „genau richtig“, von 24 Prozent als „zu lang“ und von 4 Prozent als „zu kurz“ bewertet wurde. Für den Großteil der Benutzer*innen hatten die Texte somit die richtige Länge. Das könnte einen Hinweis darauf geben, dass die Schätzung des Wissensstands der Benutzer*innen und die Adaption der Lerntexte durch das System das gewünschte Ergebnis erzielte. Für immerhin 24 Prozent waren die Texte jedoch zu lang. Für diese Gruppe sollten möglicherweise die Texte optimiert und in der Gesamtlänge gekürzt als auch eine feinstufigere Adaptivität der Texte eingeführt werden.[5] Die Textverständlichkeit wurde ebenfalls positiv, mit einer Durchschnittsschulnote von 2,1, bewertet.[6]

Bemerkenswert war zudem, dass keine signifikante Korrelation zwischen dem Alter bzw. der Vorerfahrung und dem Textverständnis, sowie zwischen dem Alter bzw. der Vorerfahrung und der Gesamtsumme an gewonnenen Trophäen festgestellt werden konnte. Die Texte scheinen sich daher für alle Benutzer*innen gleichermaßen zu eignen, was an der Strukturierung der Texte sowie deren Adaptierung an den Wissensstand der Benutzer*innen liegen könnte. Somit wird eine gleichbleibende Textverständlichkeit bei sinkendem Erfahrungswert durch einen steigenden Detailierungsgrad und einer steigenden Länge des Textes erreicht.

[5] Mit den hier erfassten Variablen lässt sich diese Gruppe nicht eindeutig klassifizieren. Um eine bessere Adaptivität anbieten zu können, müssten daher zukünftig die Eigenschaften dieser Benutzer*innen genauer erfasst werden.

[6] 20 Prozent der Benutzer*innen haben die Textverständlichkeit mit der Schulnote „sehr gut“, 62 Prozent mit der Schulnote „gut“ bewertet. 10 Prozent empfanden die Textverständlichkeit als „befriedigend“. Jeweils 4 Prozent gaben die Note „ausreichend“ und „mangelhaft“.

### 4.2 Motivation, Kreativität und Spaß

Motivation und Spaß sind wichtige Faktoren für den Lernerfolg. Auch die Möglichkeit, Probleme kreativ lösen zu können, ist eine Bedingung für erfolgreiches Lernen. In PLAY & CODE sollen diese Faktoren vor allem durch den Einsatz von Serious Games und Trophäen als Gamification-Element erzeugt werden. Der Erfolg dieser Komponenten wurde daher überprüft. Außerdem wurde untersucht, wie schwierig die Schüler*innen die Aufgaben empfanden und wie sie die Relevanz des Gelernten bewerten.

Nach dem Spaß gefragt, gaben 36 Prozent der Schüler*innen an, dass sie „sehr viel Spaß" hatten, weitere 54 Prozent hatten „viel Spaß". Nur 8 Prozent der Benutzer*innen gaben an, dass sie „wenig Spaß" hatten. Die Antwortvorgabe „überhaupt kein Spaß" wurde nicht ausgewählt. Im Durchschnitt wurde der Spaß mit 1,8 bewertet (Skala von 1 – „sehr viel Spaß" bis 4 – „überhaupt kein Spaß").[7] Bemerkenswert ist dabei die Korrelation zwischen der Dauer der Bearbeitung und dem Spaß (Einseitig, Kendalls $\tau$, $p < 0{,}05$, Korrelation: -0,295).[8] Anders als zunächst angenommen, nahm der Spaß mit steigender Dauer nicht ab, sondern zu. Dies könnte einen Hinweis darauf geben, dass die Schüler*innen daran interessiert waren, die Level, trotz der damit verbundenen Herausforderung, zu lösen. Vermutlich sind diese Effekte aber bis zu einer gewissen Dauer begrenzt und eine Frustration stellt sich danach dennoch ein.

Gefragt nach der Motivation durch Trophäen, gaben 26 Prozent der Schüler*innen „sehr motivierend" und 46 Prozent der Schüler*innen „motivierend" an. 22 Prozent der Schüler*innen fanden die Trophäen „wenig motivierend", 6 Prozent „überhaupt nicht motivierend". Im Durchschnitt empfanden die Schüler*innen die Trophäen als „motivierend" (2,3 bei Bewertungsskala 1 bis 4) mit einer Standardabweichung von 1,2. Die lediglich „gute" Bewertung verwundert insofern, als dass 90 Prozent aller Teilnehmenden mehrmals pro Woche oder häufiger Videospiele spielen und daher angenommen wurde, das Gamification-Elemente, wie sie auch in Videospielen vorkommen, mit einer deutlich positiveren Rückmeldung der Benutzer*innen verbunden ist. Möglicherweise war der unmittelbare Nutzen der Trophäen nicht klar und deutlich ersichtlich. Dieser Zusammenhang sollte zukünftig weiter untersucht werden.

Mit dem Spaßempfinden korreliert auch der Wissenszuwachs. Schüler*innen, die selbst einschätzen, viel dazugelernt zu haben, hatten nach eigenen Aussagen auch mehr Spaß beim Spiel (Skala 1 – "sehr viel Spaß" bis 5 „überhaupt kein Spaß", Einseitig, Pearson, $p < 0{,}05$, Korrelation: -0,38). Dies gibt einen Hinweis darauf, dass sich die von Keller [Ke87] beschriebene Satisfaction eingestellt hat: Durch sichtbare Erfolge wurde mehr Spaß empfunden.

Um die Kreativität der Benutzer*innen bei der Lösung der Aufgaben zu ermitteln, wurden Levenshtein-Distanzen zwischen der Abgabe der Benutzer*innen und der Musterlösung

[7] Gestellt wurde die Frage „Wieviel Spaß hat Dir das Lernen mit Spielen gemacht?" mit folgenden Antwortmöglichkeiten „sehr viel Spaß", „viel Spaß", „wenig Spaß", „überhaupt kein Spaß".

[8] Da fünf Schüler*innen ein leeres Namenskürzel im System angegeben haben, ist hier N = 45.

für die jeweilige Aufgabe errechnet und summiert (N = 45). Je höher die Gesamtsumme, desto kreativer war der/die Benutzer*in bei der Lösung der Aufgaben. Ein Zusammenhang zwischen der Kreativität der Benutzer*innen und dem Spaß konnte zwar nicht festgestellt werden, jedoch hängen Kreativität und die jeweilige Ergebnissumme zusammen: Je kreativer die Schüler*innen bei der Lösung ihrer Aufgaben waren, desto bessere Ergebnisse konnten sie erzielen (Pearson, $p < 0{,}01$, Korrelation: 0,42). Insbesondere bei Aufgaben mit optionalen Zielen kreierten viele Schüler*innen individuelle Lösungen, welche ebenfalls zum bestmöglichen Ergebnis führten. Der Schwierigkeitsgrad der Aufgaben wurde von 80 Prozent der Schüler*innen als „genau richtig" bewertet. 8 Prozent fanden die Aufgaben „zu leicht", 12 Prozent „zu schwer". Die Aufgaben waren im Sinne Kellers [Ke87] „Confidence" demnach fordernd, aber machbar.

### 4.3 Lernerfolg

Nachdem im vorherigen Abschnitt die Faktoren untersucht wurden, die einen Einfluss auf den Lernerfolg haben, soll nun der Lernerfolg der Schüler*innen selbst untersucht werden. Hierzu werden die Ergebnisse des Posttest-Quiz näher betrachtet und die Frage gestellt, ob die Schüler*innen das Erlernte im Posttest-Quiz anwenden konnten.

Der Wissenszuwachs wurde als die Differenz zwischen dem Pretest-Ergebnis und dem Posttest-Ergebnis berechnet. Die Ergebnisse der Tests konnten jeweils Werte zwischen 0 und 8 annehmen. Wurden 8 Punkte erreicht, konnten jeweils mehr als die Hälfte der Fragen der acht Themenbereichen korrekt beantwortet werden. Bei 0 Punkten wurde kein Themenbereich korrekt beantwortet. Angestrebt wurde eine möglichst große Differenz als Wissenszuwachs zwischen dem Pretest- und dem Posttest-Ergebnis.

Im Durchschnitt konnten die Schüler*innen im Posttest 4,2 Fragenkategorien mehr als im Pretest korrekt beantworten und damit ihr Ergebnis des Pretests um 50 Prozent verbessern.[9] Dabei wurden insbesondere Fragen zu Kommentaren, Vergleichsoperatoren, Datentypen und If-Abfragen richtig beantwortet.[10] Die Schüler*innen haben somit eine grundsätzliche Vorstellung von Programmierung und Wissen um die Datentypen und konditionalen Verzweigungen in JavaScript durch das Spiel erlangt.

[9] Die Evaluierung fand pro Klasse in einer Schuldoppelstunde (90 Minuten) statt. Die Schüler*innen konnten in dieser Zeit durchschnittlich 10 von 12 Lektionen mit einer Standardabweichung von 2,043 absolvieren (N = 45). Da die 12. Lektion eine zusammenfassende Aufgabe darstellt, haben die Schüler*innen im Durchschnitt zehn der elf Themenbereiche bearbeitet, lediglich for-Schleifen (Lektion 11) wurden von 60 Prozent der Teilnehmenden nicht gelernt.

[10] Kommentare: 40 korrekte und 10 falsche Antworten; Vergleichsoperatoren: 40 korrekte und 10 falsche Antworten; Datentypen: 43 korrekte und 7 falsche Antworten; If-Abfragen: 40 korrekte und 10 falsche Antworten; Variablen: 24 korrekte und 26 falsche Antworten; for-Schleifen: 16 korrekte und 34 falsche Antworten; Funktionen: 21 korrekte und 29 falsche Antworten; Logik-Operatoren: 11 korrekte und 39 falsche Antworten.

## 5 Zusammenfassung und Fazit

Die gesellschaftliche Bedeutung der Informatik nimmt kontinuierlich zu, gleichzeitig fehlt es an Fachkräften. Vielen erscheint Programmierung als zu abstrakt und mühsam. Es müssen deshalb neue Wege gefunden werden, Programmieranfänger*innen zu motivieren und das notwendige Fachwissen praxisnäher zu vermitteln. Es stellte sich daher die Frage, ob mithilfe eines Serious Games basierten Lernsystems Motivation und Lernerfolg für grundlegende Programmierkenntnisse gesteigert werden können. Zur Beantwortung dieser Frage wurde die Serious Games basierte Lernplattform PLAY & CODE entwickelt.

Um die Wirksamkeit dieser Lernplattform zu überprüfen, wurde eine Evaluation an mehreren Schulen in Berlin durchgeführt. Den Schüler*innen wurden mit der Lernplattform Grundlagen von JavaScript vermittelt. Dabei wurde gezeigt, dass die Schüler*innen bei der Lösung der Aufgaben motiviert waren und Spaß hatten. Mittels eines Pre- und Posttests wurde festgestellt, dass die Schüler*innen nach dem Kurs 50 Prozent mehr Fragen zu den gelehrten Themenbereichen korrekt beantworten konnten als davor. Die Schüler*innen hatten nach dem Spiel eine genauere Vorstellung von Programmierung und verfügten über einige Grundlagen von JavaScript. Die Forschungsfrage kann daher positiv beantwortet werden: Serious Games basierte Lernsysteme wie PLAY & CODE sind ein wirkungsvolles Mittel zur Vermittlung von Programmiergrundlagen.

Um die Qualität der Plattform noch zu steigern, wird in Zukunft eine feinstufigere Adaptivität der Texte in Erwägung gezogen. Untersucht werden sollte auch, wie die Motivation durch Trophäen weiter gesteigert werden kann. Auch bei den Erhebungsinstrumenten gibt es Optimierungsbedarf. Zukünftig sollten in den Fragebögen die Namensfelder erzwungen werden, um eine höhere Anzahl an vollständigen Datensätzen zu erzielen. Außerdem werden in kommenden Evaluierungen im Pre- und Posttest jeweils die gleiche Anzahl an Fragen pro Themenbereich gestellt. Damit kann der Wissenszuwachs genauer bestimmt werden. Zudem wurde die Lernplattform ausschließlich in Informatik-Schulklassen evaluiert. Da angenommen werden kann, dass diese Schüler*innen ein gesteigertes Verständnis für Programmierung haben, wurde das Ergebnis dadurch womöglich verzerrt. PLAY & CODE sollte daher zukünftig auch von Personen aus dem Hobby- und Selbstlernbereich evaluiert werden.

## 6 Danksagung

Ich danke Herrn Prof. Dr. Pinkwart für seine wertvollen Kommentare und Anmerkungen bei der Erstellung dieses Beitrags.

Literaturverzeichnis

[Ch15] Cheng, M. et. al.: The use of serious games in science education: a review of selected empirical research from 2002 to 2013. Journal of Computers in Education, 02/15: S. 353-375, 2015.

[Dö16] Dörner, R.; Göbel, S.; Effelsberg, W.; Wiemeyer, J.: Serious Games. Foundations, Concepts and Practice, Springer International Publishing, Basel, 2016.

[GEM13] Girard, C.; Ecalle, J.; Magnan, A.: Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. Journal of Computer Assisted Learning, 29/13: S. 207-219, 2012.

[He17] Heublein, U. et. al.: Motive und Ursachen des Studienabbruchs an baden-württembergischen Hochschulen und beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher. DZHW Projektbericht, 6/17: 2017.

[Hu07] Hubwieser, P.: Didaktik der Informatik. Grundlagen, Konzepte, Beispiele, 3. Auflage, Springer-Verlag Berlin Heidelberg, Berlin, 2007.

[IW17] IW, Institut der deutschen Wirtschaft, https://www.iwkoeln.de/presse/iw-nachrichten/beitrag/oliver-koppel-mint-fachkraefte-verzweifelt-gesucht-350061.html, Stand: 09.03.2019

[Ka12] Kazimoglu, C. et. al.: A Serious Game for Developing Computational Thinking and Learning Introductory Computer Programming. Procedia - Social and Behavioral Sciences, 47/12: S. 1991-1999, 2012.

[Ke87] Keller, J. M.: Development and use of the ARCS model of instructional design. Journal of instructional development, 3/87, S. 2-10, 1987

[Me02] Merrill, M. D.: First principles of instruction, Educational technology research and development, 3/02: S. 43-59, 2002.

[Mo14] Morgan, N.: JavaScript for Kids. A Playful Introduction to Programming, No Starch Press, San Francisco, 2014.

[Mu09] Muratet, M. et. al.: Towards a serious game to help students learn computer programming. International Journal of Computer Games Technology, 3/09: S. 1-12, 2009.

[Sa06] Savery, J. R.: Overview of problem-based learning: Definitions and distinctions. Interdisciplinary Journal of Problem-Based Learning, 1/06: S. 9-20, 2006.

[Sc13] Schäfer, A. et. al.: M.: From boring to scoring. a collaborative serious game for learning and practicing mathematical logic for computer science education. Computer Science Education, 23/13: S. 87-111, 2013.

[TI19] TIOBE Programming Community Index, https://www.tiobe.com/tiobe-index/, Stand: 09.03.2019

[VDC11] Vandewaetere, M.; Desmet, P.; Clarebout, G.: The contribution of learner characteristics in the development of computer-based adaptive learning environments. Computers in Human Behavior, 27/11: S. 118-130, 2011.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 247

# Die Self-Determination Theory als Rahmen für Deep Gamification in didaktischen Kontexten

Heinrich Söbke[1]

**Abstract:** Unter Gamification wird die Anwendung von Spielprinzipien auf reale Kontexte verstanden. Der Begriff Deep Gamification wird verwendet, wenn die Prozesse realer Kontexte strukturell durch Gamification verändert werden. Deep Gamification führt dazu, dass die Nutzer in die Lage versetzt werden, die Prozesse gemäß persönlichen Präferenzen anzupassen. Im Vergleich zu Gamification wird davon ausgegangen, dass Deep Gamification eine stärkere Wirkung auf die Motivation der Nutzer hat. Bislang gibt es jedoch keinen umfassenden konstruktiven Rahmen – im Sinne von Gestaltungsrichtlinien – für die Integration von Deep Gamification in didaktische Kontexte. Dieser Artikel untersucht die Self-Determination Theory (SDT) als Grundlage eines konstruktiven Rahmens für die Integration von Deep Gamification in didaktische Kontexte. Eine Fallstudie, die außergewöhnlich hohe Lernaktivitäten der Studierenden zeigt, weist Merkmale von Deep Gamification auf und dient als Ausgangsbasis. Die Anwendbarkeit der SDT wird durch einen Abgleich mit den Gamification-Elementen der Fallstudie gezeigt.

**Keywords:** Deep Gamification, SDT, Hochschulbildung, Motivations-Design, Gamification

## 1 Einleitung

Unter Gamification wird die Anwendung von Spielelementen auf Nicht-Spiel-Kontexte der realen Welt verstanden [De11]. Das Ziel von Gamification ist es, über eine gesteigerte Motivation die Aktivitäten der Nutzer in den gamifizierten Kontexten zu erhöhen. In den letzten Jahren gab es verschiedene Ansätze, Gamification in didaktischen Kontexten – einem Anwendungsfeld von Gamification – zu nutzen, wie der Überblick von Majuri et al. verdeutlicht [MKH18]. Die Kategorisierung der genutzten Gamification-Elemente zeigt, dass die Mehrheit der Gamification-Elemente individuelle Leistung und Fortschritt honoriert, wie beispielsweise Punkte, Abzeichen und Ranglisten. Nur wenige Studien untersuchen Gamification-Elemente, die auf sozialer Ebene wirken, beispielsweise Kooperation oder Peer-Review fördern. Fast 30 % der Studien identifizieren auch negative Wirkungen von Gamification. Gleichzeitig berichten Subhash und Cudney von einer zunehmenden Nutzung von Gamification in der Hochschulbildung [SC18]. Der Einsatz von Bildungstechnologie fördert die Verbreitung von Gamification als integralen Bestandteil der Bildungstechnologie. Dementsprechend ist Gamification aktueller Gegenstand der Forschung [ND17]. Mekler et al. stellen fest, dass Gamification-Elemente einen Einfluss auf die Aktivitäten haben, jedoch nicht die Wahrnehmung von Kompetenz und intrinsischer Motivation ändern

[1] Bauhaus-Universität Weimar, Bauhaus-Institut für zukunftsfähige Infrastruktursysteme, Coudraystr. 7, 99423 Weimar, heinrich.soebke@uni-weimar.de

doi:10.18420/delfi2019_334

[Me17]. Die untersuchten Gamification-Elemente, wie Punkte, Abzeichen und Ranglisten, sind daher als extrinsisches Motivationselement zu betrachten. Andere Studien beobachten weitergehend, dass die extrinsisch wirksamen Gamification-Elemente wie Punkte, Abzeichen und Ranglisten zu einer Reduzierung der intrinsischen Motivation führen [DID18], [HF15], [HKS14]. Eigene Untersuchungen deuten an, dass die Effekte von Gamification auch durch die aufwandsoptimierende Zielorientierung der Lernenden gemindert werden [SR18]. Es gibt verschiedene Ansätze, um den beschriebenen Defiziten von Gamification zu begegnen, wie adaptive Gamification [BNB17]. Ein weiterer Ansatz ist unter der Bezeichnung *Deep Gamification* bekannt. Eine pragmatische Definition von Deep Gamification in didaktischen Kontexten liefert Santos [Sa15]: das wesentliche Kriterium von Deep Gamification ist die strukturelle Änderung der Prozesse des zu gamifizierenden Kontexts.

## 2 Methodik

Eine eigene in Bezug auf Aktivitätssteigerung erfolgreiche Fallstudie dokumentiert das enorme Potenzial von Deep Gamification [SZ18]. Der didaktische Kontext wurde mit dem Ziel des Motivationsdesigns entworfen, Gamification ist hier eher als Nebenprodukt des Motivationsdesigns zu sehen. Es stellt sich die Frage, wie Deep Gamification gezielt in einen didaktischen Kontext integriert werden kann. Ziel der Arbeit ist es, einen konstruktiven Rahmen zu beschreiben, auf dessen Grundlage systematisch Gestaltungsrichtlinien für die Integration von Deep Gamification in didaktische Kontexte erstellt werden können. Zur Identifikation eines konstruktiven Rahmens wird ein Theoriemodell betrachtet, das mit den Gamification-Elementen der Fallstudie abgeglichen wird und als Grundlage eines konstruktiven Rahmens gelten kann. Dahinter steht die Überlegung, dass die Fallstudie lediglich einen Teil möglicher Gestaltungsrichtlinien umsetzt, ein Theoriemodell hingegen als vollständig anzusehen ist. Wenn nun ein Theoriemodell als valide Grundlage von Deep Gamification identifiziert werden kann, dann lässt sich aus dem Theoriemodell systematisch eine – wenn auch nicht vollständige, dann zumindest umfassendere – Menge von Gestaltungsrichtlinien ableiten. Als Theoriemodell wurde die Self-Determination Theory (SDT) gewählt [RD17]. Die SDT ist eine allgemeine Motivationstheorie, die in Zusammenhang mit Serious Games und Gamification häufig angewendet wird (z.B. [Ka12], [LP18]).

## 3 Fallstudie: Eine Quiz-App als fakultatives Lernwerkzeug

Der Kurs Siedlungswasserwirtschaft des Bachelorstudiengangs Bauingenieurwesen erfordert von den Studierenden, zunächst Faktenwissen zu erwerben, dass dann als Grundlage für prozedurales Wissen, wie beispielsweise die Dimensionierung einer Kläranlage, genutzt werden kann. Der Erwerb von Faktenwissen wird durch eine Reihe von Zulassungstests angeregt, die im Verlaufe des Semesters bestanden werden müssen,

um die Zulassung zur Abschlussprüfung zu erreichen. Jeder der über das Learning Management System (LMS) Moodle angebotenen Zulassungstests besteht aus fünf Multiple-Choice-Fragen, konzentriert sich auf ein Vorlesungsthema und muss innerhalb von 14 Tagen nach der Vorlesung bestanden werden. Zulassungstests können zusätzlich mit einer Quiz-App vorbereitet werden. Die in der Quiz-App 10 % bestplatzierten Studierenden (Ranglistenkriterium: korrekt beantwortete Fragen) müssen den Zulassungstest nicht ablegen – sie erhalten ein sogenanntes "Freiticket". Darüber hinaus werden je nach Anzahl der gespielten Partien aller Kursteilnehmer zusätzliche Beispielklausuren für den gesamten Kurs als Gruppenanreiz angeboten. Die Studie im Wintersemester 2018 zeigte eine sehr hohe Aktivität: mehr als 2000 Fragen wurden von jedem Studierenden im Durchschnitt beantwortet [SZ18]. Eine solch hohe Aktivität übersteigt bei weitem den Aktivitätsgrad früherer Quiz-App-Studien im gleichen Kurs [SR18].

## 4 Self-Determination Theory

Eine häufig verwendete theoretische Grundlage zur Analyse von Gamification und Serious Games aus psychologischer Sicht ist die SDT [LB18]. So beschreiben Dicheva et al. einen konstruktiven, auf der SDT basierenden Entwurfsansatz für ein gamifiziertes LMS [DID18]. Die SDT identifiziert drei grundlegende psychologische Bedürfnisse des Menschen – (I) Autonomie, (II) Kompetenz und (III) Soziale Eingebundenheit – als Treiber der intrinsischen Motivation. Autonomie kann kurz definiert werden als Handeln nach eigenen Entscheidungen, sollte aber nicht mit der Unabhängigkeit von anderen Personen und Umständen verwechselt werden. Unter Kompetenz wird die Fähigkeit verstanden, angestrebte Ergebnisse auch zu erreichen und als wichtig erachtete Vorgänge im eigenen Sinne beeinflussen zu können. Soziale Eingebundenheit kann als der Wille gesehen werden, mit anderen zu interagieren und Teil einer Gruppe zu sein [RD17]. Loughry & O´ Broin [LB18] geben einen detaillierten Überblick über den Stand der Forschung zur SDT als theoretische Grundlage von Gamification. Sie kommen zu dem Schluss, dass die SDT als eine gültige theoretische Grundlage von Gamification angesehen werden kann. Darüber hinaus folgern sie, dass die Bedeutung der Belohnungen zur Motivation beiträgt: Belohnungen mit hoher funktionaler Bedeutung wirken als extrinsische Motivatoren und erhöhen die intrinsische Motivation. Sie empfehlen jedoch, extrinsische Motivatoren vorsichtig einzusetzen.

## 5 Die Fallstudie aus Sicht der SDT

Dieser Abschnitt betrachtet die Fallstudie aus der Sicht der SDT. Ziel ist es, SDT-basierte Charakteristiken der Fallstudie zu identifizieren. Die identifizierten Charakteristiken können als Beispiele der Umsetzung der SDT in einem didaktischen Kontext gesehen werden, eine Verallgemeinerung kann dann als Gestaltungsrichtlinie für didaktische Kontexten gemäß der SDT genutzt werden und damit Deep Gamification

unterstützen. Ein Rahmen für SDT-bezogenes Design wird von Shi und Cristea beschrieben [SC16]. Allerdings ist das Rahmenwerk von Shi und Cristea spezifisch auf E-Learning-Kontexte ausgerichtet, während die Fallstudie Teil eines Präsenzstudiums ist. Im Folgenden werden die Fallstudie und die Richtlinien, die sogenannten Gamification-Strategien des Rahmens von Shi und Cristea im Sinne einer Validierung am Beispiel der Richtlinien für Autonomie abgeglichen. Insgesamt konnten alle von Shi und Christea genannten Richtlinien mit Charakteristiken der Fallstudie zur Deckung gebracht werden.

**A1. Eine Menge von Lernzielen mit klaren Beschreibungen und mehreren Möglichkeiten, diese zu erreichen.** Es gibt es verschiedene Möglichkeiten, für die Abschlussprüfung zugelassen zu werden. Für jedes Vorlesungsthema kann der Studierende entweder den Zulassungstest ablegen oder eine Mindestanzahl von Fragen in der Quiz-App beantworten. Darüber hinaus kann ein Zulassungstest "bestanden" werden, indem der Studierende in Bezug auf die Anzahl der beantworteten themenspezifischen Fragen unter den besten 10 % der Teilnehmer liegt. Die Alternativen erfordern jeweils einen guten Wissensstand des zu lernenden Faktenwissens.

**A2. Unterschiedliche Werkzeuge zur Erledigung einer Aufgabe.** Die Fallstudie verwendet eine Quiz-App und Zulassungstests als alternative Werkzeuge, um die Zulassung zur Abschlussprüfung zu erreichen.

**A3. Klare, sofortige und positive Rückmeldungen zu Lernaktivitäten.** Sowohl die Quiz-App als auch die Zulassungstests geben sofortige Rückmeldungen. Die Rückmeldungen sind jedoch nur dann positiv, wenn eine Frage richtig beantwortet wird oder ein Zulassungstest bestanden wird. Hier fehlen noch ermutigende Kommentare.

**A4. Optionen mit Konsequenzen.** Beide Optionen (Zulassungstest oder Quiz-App) erlauben einzeln, das Ziel der Zulassung zur Abschlussprüfung zu erreichen.

**A5. Lernkontext, der von den Lernenden selbst angepasst werden kann.** Die Studierenden können gezielt Zulassungstests ausschließen, indem sie entsprechende „Freitickets“ sammeln.

## 6 Diskussion

Für die SDT konnten in jeweils allen Kategorien Charakteristiken der Fallstudie zur Deckung gebracht werden. Damit ergeben sich erste Gestaltungsrichtlinien. In einem weiteren, über diese Arbeit hinausgehenden, Schritt ist es notwendig, aus den gefundenen Charakteristiken allgemeingültige Gestaltungsrichtlinien abzuleiten, die einen konstruktiven Rahmen für die Integration von Deep Gamification in didaktische Kontexte bilden. Unter Umständen weisen neben der SDT weitere Theoriemodelle eine bessere Eignung auf oder können als Ergänzung dienen, beispielsweise die Spielkomponenten nach Caillois [Ca01]. Eine andere Aufgabe ist die systematische Integration der bestehenden Taxonomien von Gamification-Elementen (z.B. in

[MKH18]) in die Gestaltungsrichtlinien. Als zusätzliche Quellen zur Ergänzung der Gestaltungsrichtlinien können bereits bestehende Richtlinien für die Integration herkömmlicher Gamification-Elemente in didaktische Kontexte (z.B. [BH18]) dienen.

## 7 Zusammenfassung

Als Anlass dieser Arbeit wurde eine außergewöhnlich hohe Aktivität von Studierenden in einem didaktischen Kontext auf Deep Gamification zurückgeführt. Während Gamification als die Anwendung von Spielprinzipien auf Kontexte der realen Welt zu sehen ist, liegt Deep Gamification dann vor, wenn Gamification eines Kontextes zu einer Änderung der zugrundeliegenden Prozesse führt: abhängig von den Entscheidungen der Nutzer existieren individuelle Wege durch die Prozesse. Deep Gamification wurde in dieser Arbeit mit der Self-Determination Theory (SDT) abgeglichen. Wichtige Merkmale von Deep Gamification sind aus Sicht der SDT die Bereitstellung von Entscheidungsmöglichkeiten für die Nutzer, kurze Zyklen von Aktionen und sofortigen Rückmeldungen sowie die Förderung sozialer Kontexte für die Nutzer.

Literaturverzeichnis

[BH18] Brunvand, S.; Hill, D.: Gamifying your Teaching. Guidelines for Integrating Gameful Learning in the Classroom. In College Teaching, 2018, 0; S. 1–11.

[BNB17] Böckle, M.; Novak, J.; Bick, M.: Towards Adaptive Gamification. A Synthesis of Current Developments: Conference: Twenty-Fifth European Conference on Information Systems (ECIS), Guimarães, Portugal, 2017.

[Ca01] Caillois, R.: Man, Play and Games. University of Illinois Press, 2001.

[De11] Deterding, S. et al.: From game design elements to gamefulness. Defining gamification: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. ACM, New York, 2011; S. 9–15.

[DID18] Dicheva, D.; Irwin, K.; Dichev, C.: OneUp. Supporting Practical and Experimental Gamification of Learning. In International Journal of Serious Games, 2018, 5; S. 5–21.

[HF15] Hanus, M. D.; Fox, J.: Assessing the Effects of Gamification in the Classroom. In Computers and Education, 2015, 80; S. 152–161.

[HKS14] Hamari, J.; Koivisto, J.; Sarsa, H.: Does gamification work? - A literature review of empirical studies on gamification: Proceedings of the Annual Hawaii International Conference on System Sciences, 2014; S. 3025–3034.

[Ka12] Kapp, K. M.: The Gamification of Learning and Instruction. Game-based Methods and Strategies for Training and Education. Pfeiffer, San Francisco, 2012.

[LB18] Loughrey, K.; Broin, D.: Are We Having Fun Yet? Misapplying Motivation to Gamification: 2018 IEEE Games, Entertainment, Media Conference (GEM), 2018; S. 1–9.

[LP18] Lamprinou, D.; Paraskeva, F.: Gamification design framework based on SDT for student motivation: 2018 IEEE Games, Entertainment, Media Conference (GEM), 2018; S. 406–410.

[Me17] Mekler, E. D. et al.: Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. In Computers in Human Behavior, 2017, 71; S. 525–534.

[MKH18] Majuri, J.; Koivisto, J.; Hamari, J.: Gamification of education and learning. A review of empirical literature: CEUR Workshop Proceedings, 2018; S. 11–19.

[ND17] Nacke, L. E.; Deterding, C. S.: The maturing of gamification research. In Computers in Human Behavior, 2017; S. 450–454.

[RD17] Ryan, R. M.; Deci, E. L.: Self-determination theory. Basic psychological needs in motivation, development, and wellness. Guilford Publications, 2017.

[Sa15] Santos, P. A.: Deep Gamification of a University Course: Conference Proceedings SciTecIN'15 Sciences and Technologies of Interaction 12-13 November 2015 Coimbra, Portugal. Departamento de Engenharia Electrotécnica Polo II da Universidade de Coimbra, 2015.

[SC16] Shi, L.; Cristea, A. I.: Motivational gamification strategies rooted in self-determination theory for social adaptive E-learning: International Conference on Intelligent Tutoring Systems. Springer, 2016; S. 294–300.

[SC18] Subhash, S.; Cudney, E. A.: Gamified learning in higher education. A systematic review of the literature. In Computers in Human Behavior, 2018, 87; S. 192–206.

[SR18] Söbke, H.; Reichelt, M.: Sewer Rats in Teaching Action: An explorative field study on students' perception of a game-based learning app in graduate engineering education: CoRR, abs/1811.0, 2018; Abgerufen von http://arxiv.org/abs/1811.09776

[SZ18] Söbke, H.; Zander, S.: Motivationsdesign durch Verschränkung von Gamifikation und didaktischem Kontext: Eine Quiz-App in einem ingenieurtechnischen Studiengang. In (Krömker, D.; Schroeder, U., (eds.); DeLFI 2018 - Die 16. E-Learning Fachtagung Informatik. Gesellschaft für Informatik e.V., Bonn, S. 141–152, 2018.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 253

# Interaktive Unterstützung bei der Begehung eines Kunstlehrpfads: Die WaldSkulpturenWeg-App

Sven-Bjarne Seiffert[1] und Andreas Lingnau[2]

**Abstract:** Museumsexponate einem Besucher mit Hilfe einer App näher zu bringen ist durch die Verfügbarkeit von Smartphones eine Alternative zu Audioguides oder stationären Monitoren. Die Annahme, dass die meisten Besucher im Besitz eines Smartphones sind, erlaubt die Entwicklung neuer Konzepte für museumsspezifische Software unter Zuhilfenahme technischer Möglichkeiten wie z.B. Ortungsdienste und Sensoren in Smartphones. Wir beschreiben die WaldSkulpturenWeg-App, deren Ziel es ist, ein umfangreiches Informationsportal zu schaffen, welches den Besuchern im Rahmen eines Kunstprojektes im Außenbereich einen interaktiven Umgang mit Kunstwerken ermöglichen soll und Informationen, die im Zusammenhang zum Kunstprojekt stehen und auf die Besichtigung der Skulpturen ausgerichtet sind, zu vermitteln. Darüber hinaus schafft ein integriertes Gamification-System einen zusätzlichen Anreiz, mehr über die Kunstwerke zu lernen.

**Keywords:** Museumsdidaktik, Gamification, informelles Lernen.

## 1 Einleitung

Geographisch durch den Kamm des Rothaargebirges getrennt und kulturhistorisch – insbesondere durch die Reformation – unterschiedlich geprägt, entstand zwischen den Gemeinden Bad Berleburg und Schmallenberg im Zeitraum von zehn Jahren, beginnend im Jahr 2000, eine Art kultureller Brückenschlag: der WaldSkulpturenWeg (WSW). Unter Beteiligung international bekannter Künstler wurden auf der 23 km langen Wanderroute 11 Kunstwerke am Wegesrand platziert. Der Weg ist damit sowohl ein Wanderpfad als auch ein Museum, der eine Vielzahl von Touristen in die Region lockt.

Museumsexponate einem Besucher mit Hilfe einer zugeschnittenen App näher zu bringen, ist seit der umfassenden Verfügbarkeit von Smartphones eine Alternative zu eigenständigen Gerätegruppen wie mobilen Audioguides und stationären Monitoren. Die Annahme, dass die meisten Besucher im Besitz eines Smartphones sind, schafft neue Möglichkeiten, da diese über Ortungsdienste und Sensoren verfügen, die tragend für museumsspezifische Software sein können. Im Rahmen einer Masterarbeit wurde die WaldSkulpturenWeg-App entworfen, deren Konzept darauf abzielt, ein umfangreiches Informationsportal zu schaffen, das den Besuch der Anlage zu einem eindrucksvolleren Erlebnis machen soll. So soll die Orientierung durch Karten, die für den WSW zugeschnitten wurden, stark erhöht werden. Besucher können nicht nur ihre Position verfolgen, wie es in herkömmliche Karten-Apps möglich ist, sondern erhalten Wegbezogene Informationen, die direkt im Zusammenhang zum Kunstprojekt stehen und auf die Besichtigung der Skulpturen ausgerichtet sind. Jedes Kunstwerk steckt voller Informa-

[1] Humboldt-Universität zu Berlin, Institut für Informatik, Berlin
[2] Hochschule Ruhr West, Bottrop

doi:10.18420/delfi2019_221

tionen, die dem Besucher über eine App auf anderen Wegen als auf statischen Tafeln vermittelt werden können und ihn so motivieren, sich interaktiv mit den Kunstwerken auseinanderzusetzen. GPS-Daten können die Informationsvermittlung steuern oder ankündigen und in Form von Audio, Text und Bildern übertragen. Ein Anreiz für einen längeren Aufenthalt auf dem WSW schafft die Integrtation eines Gamification-Systems.

Das Modell der WSW-App lässt sich in zwei Kategorien einordnen: Zum einen stellt die App ein Informationsportal für Ausstellungsstücke dar und ist somit eine Museums-App, zum anderen finden sich aufgrund der Beschaffenheit des Kunstprojektes viele Lokalisierungsdienste wieder, was die Anwendung so auch zu einer Navigations-App macht. Die App des Deutschen Technik Museums[3] zeigt Parallelen zum Konzept der WSW-App. Sie simuliert einen klassischen Audioguide und sie bietet einen spielerischen Teil mit einem Quiz zu den Ausstellungsstücken inklusive einer virtuellen Urkunde. Aufgrund der funktionalen Beschränkung ist sie einfach zu bedienen, bietet aber keine weitergehenden Informationen zu Ausstellungsinhalten. Eine Motivation, das Museum zu besuchen wird dadurch nur bedingt erzeugt. Die Weitläufigkeit der Exponate beim WSW und die Lokalisierungsfunktion zeigen Analogien zur Stolpersteine-App und dem Kunstprojekt „Stolpersteine“ [CO14]. Im Gedenken an die Opfer des Nationalsozialismus wurden kleine Messingtafeln vor deren ehemaligen Wohnorten verlegt, von denen in der App eine Auswahl auf einer Karte gezeigt wird. Durch Antippen werden Informationen zu den Opfern abgerufen. Zu bestimmten Gruppierungen von Steinen gibt es Routenvorschläge. Das Konzept der App ist die Navigation durch die Zusatzinformationen, eine inhaltshinterlegte Auseinandersetzung zur Begründung und Motivation dieses Kunstprojektes findet allerdings nicht statt. Hier knüpft das Konzept der WSW-App an: Um die Auseinandersetzung mit der Kunst zu steigern, bedarf es einer zusätzlichen Präsentation von Inhalten, die durch Betrachtung oder lesen eines Informationstextes nicht ersichtlich werden. Die Basisinformationen der Kunstwerke werden durch Aussagen und Fragen ergänzt, die positionsspezifische Inhalte aufweisen können. Dazugehörige Texte werden in Hinweispunkten auf der Landkarte untergebracht und bieten dem Besucher so die Möglichkeit, aus bestimmten Blickwinkeln und Positionen im Skulpturenumfeld einer bestimmten Anregung zu folgen.

## 2 Theoretisch-Didaktische Hintergründe

Im Folgenden werden Zweck und Zielsetzung der App im Hinblick auf Vermittlungstheorien in Museen sowie des Gamification-Konzepts erläutert und erste Ergebnisse aus einer Pilot-Phase mit einer Besuchergruppe vorgestellt.

### 2.1 Museumspädagogik

Die moderne museale Arbeit ist sich heute einer weit umfänglicheren Form des Ausstellens bewusst, als es ihre ursprüngliche war, die als „archivalisch, konservatorischer

[3] Fluxguide. (2017) Deutsches Technikmuseum (V2.0.1).

und theoretischer Natur“ beschrieben wird [RE95]. Die Vermittlung der Inhalte ist zu einem wichtigen Teilgebiet der musealen Forschung geworden und hat sich als wissenschaftliches Forschungsgebiet gefestigt, das pädagogische und didaktische Aufgaben eines Museums untersucht. Nach [RE95] sind der gleichzeitige Gebrauch dieser sinnverwandten Wörter in der Literatur ein Problem. Er sieht in dem Ansatz, Museumsdidaktik den vermittlungsbezogenen Aufgaben in der Konzeption von Ausstellungen zuzuordnen und Museumspädagogik als personelles, praxisorientiertes Aufgabenfeld zu sehen, welches sich um Tätigkeiten wie Materialerstellung, Öffentlichkeitsarbeit und Organisation dreht, eine treffende Abgrenzung. [HE06] beschreibt *Museum Education* als handlungsorientiertes Feld, in dem Kuratoren und Ausstellungs-designer Programme entwickeln, die über die klassische Ausstellungsführung hinausgehen und zielgruppengerechte Lehrinhalte schaffen. Dies entspricht der von [RE95] bevorzugten Definition der praktischen Ausführung. So findet man auch bei [HO91], Museum Education sei die Arbeit des Museums, die Exponate für alle Gesellschafts-schichten zugänglicher zu machen. Die Arbeit beginnt dabei, bevor die Ausstellungsstücke arrangiert werden.

Die Konzeption einer Smartphone-App zum WSW-Projekt ist in der Beurteilung aus der Entwicklungsperspektive, vernehmlich mit dem Begriff der Museumspädagogik, eine Maßnahme zur höheren Zugänglichkeit der Skulpturen. Sie ordnet sich deshalb sehr gut in den Begriff ein, da sie eine methodische Umsetzung vermittlungstheoretischer Ansätze darstellt. Hauptzweck der Museumspädagogik ist es, dem Besucher zu helfen, sich gedanklich mit Hintergründen und Intentionen der Ausstellung oder des Exponats auseinanderzusetzen. Dies schließt das Verstehen einzelner Objekte sowie die Fähigkeit ein, diese in ihren Ausstellungs- oder Sammlungskontext einordnen zu können. Darüber hinaus soll eine Reflexion angeregt werden, die transitiv auch über die Grenzen des Museums hinausgehen kann. Nach [RE95] hängen die angewandten Methoden von den Gegebenheiten des Museums, dem Thema und der Zielgruppe ab. Die moderne Museumspädagogik bezieht sich zunehmend auf die Theorien allgemeiner Didaktik. So findet sich vor allem das Konzept des passiven und aktiven Lernens in den Methoden der Museumspädagogik wieder. [HE06] beschreibt das passive Lernen in Anlehnung an die klassischen Lerntheorien als ein additiv heranwachsendes Wissen, welches durch Empfangen von Informationen gekennzeichnet ist. Museen mit dieser Vermittlungs-Didaktik arbeiten mit Texttafeln an den Exponaten. Dabei ist die Ausstellung sequentiell gegliedert, mit einem klaren Anfang und Ende und einer intendierten Ordnung.

Die Vermittlung von Informationen wird in der WSW-App auf verschiedenen Wegen erbracht. Ein Textbeitrag, der zu jeder Skulptur aufrufbar ist, ähnelt einer Beschilderung, obgleich der Gebrauch eines Mobilgerätes die freie Bewegung um das Exponat erlaubt. Dies kann von „tragendem“ Vorteil sein, da es sich bei einigen Skulpturen um große, umlaufbare Objekte handelt. Ebenso lässt der Audio-Guide ein besseres Auseinandersetzen mit dem Objekt während des Hörens zu. Dem oben genannten Paradigma des „aktiven Betrachters“ soll die Anordnung von Hinweisen um die Skulpturen entsprechen. In der Detailansicht einer Skulptur können so Fragen und Anregungen in kleinen Textfeldern erscheinen, welche in die Landkarte integriert sind. Hinweise, die über bloße Informationsvermittlung hinausgehen und Besucher eher dazu anregen, sich

eigene Gedanken über das Kunstwerk zu machen, können auf den aktuellen Standort und die Blickrichtung des Nutzers auf das Objekt zugeschnitten werden.

## 2.2 Gamification

Gamification wird von [MC17; DE11] als „Anwendung von Spiel-Attributen in nichtspielerischer Umgebung“ definiert. Instrument einer Gamification-Anwendung kann z.B. die Belohnung in Form von Punkten, Abzeichen oder Bestenlisten sein. Die so entstehende Spielwelt soll den Nutzer motivieren, sich mit der Realwelt näher zu beschäftigen oder einem bestimmten Ziel zu folgen. Gamification ist jedoch kein eigenständiges Spielelement in einer Anwendung, vielmehr wird die gegebene Funktionalität zu einem Spiel erweitert. Eine Gefahr, die [MC17] erwähnt: Gamification kann manipulativ wirken, denn in welchem Maß würde der Nutzer dem Dienst ohne Gamification folgen? Nach [NI15] schafft belohnungsbasierte Gamification nur einen kurzweiligen Stimulus und bringt auf lange Sicht sogar eine Reduzierung der Motivation mit sich. Er schlägt eine „Meaningful Gamification“ vor, bei der ebenjene *Bedeutung* für den Nutzer dann entstehe, wenn er mit dem Erlebten Assoziationen zur eigenen Erfahrung herstellen kann. Sein Konzept baut auf 6 Säulen auf: Play, Exposition, Choice, Information, Engagement und Reflexion. „Entdecken“ soll durch einen spielerischen Rahmen erleichtert werden, die Darstellung soll in die Realwelt eingebunden sein, die Nutzer sollen Entscheidungen selbst treffen, die Informationsvermittlung soll spielerisch erfolgen und die Reflexion angeregt werden, um Erfahrungen und Interessen über den Rahmen hinaus zu finden, die Engagement und Lernen vertiefen können. Nach [KA14] beinhalten effektive Gamification-Umsetzungen „[...] elements of story, challenge and continual feedback as well as high level of interactivity“. [KI18] präsentieren mit Story ein weiteres Gamification-Attribut, welches sich in der WSW-App findet. Danach ist es wichtig einen Zusammenhang auftretender Spiel-Events zu kreieren, dessen Werdegang vom Nutzer initiiert wird und sich damit auch dessen Motivation steigert, die Geschichte fortzuführen.

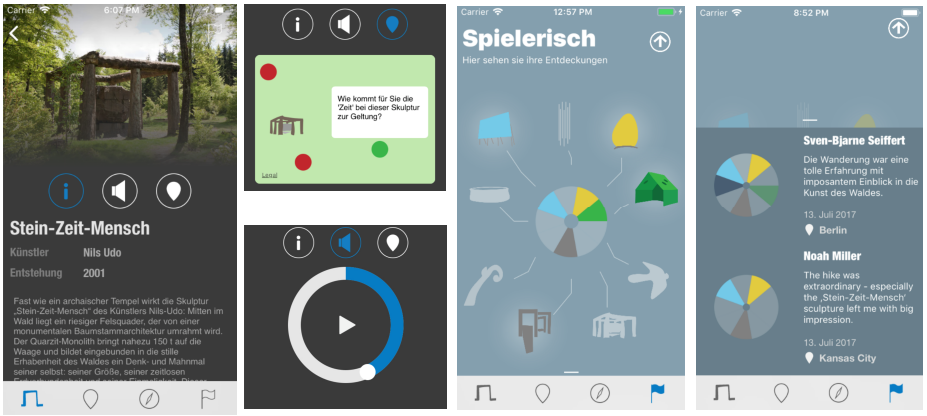



Abb. 1: Informationsportal, Karten- und Audioansicht, Gamification Komponenten.

## 3 Implementierung

Die App als Informationsportal ist in Detailansichten zu jeder Skulptur verwirklicht (Abb. 1). Die Detailansicht einer Skulptur lässt sich durch die globale Skulptur-Liste des ersten Tabs oder durch Auswählen von Elementen auf der Landkarte erreichen. Der untere Bereich der Ansicht enthält drei Kategorien, die durch Auswahl des jeweiligen Buttons erreicht werden. Hier liegt der Informationstext, der Audioguide-Player und eine abgestimmte Landkarte eines 80-Meter Umkreises der Skulptur. Auf der Landkarte sind zudem die Punkte eingetragen, die das Öffnen von Fragen und Anregungen zum Kunstwerk ermöglichen. Die Lokalisierungsdienste des Gerätes lassen durch das GPS stets die eigene Position erkennen, auch bei schlechtem Mobilfunknetz-Zugang in der Region.

Die Zielsetzungen von Gamification-Elementen werden im WSW-Konzept in zwei Bereichen umgesetzt: Ein globales Spiel bildet sich in der WSW-App durch das Belohnen von Besuchen der Skulpturen. Eine eigens dafür angelegte Seite zeigt die bisher erlaufenen Kunstwerke auf und präsentiert das Ergebnis in einer Art sich füllendem Tortendiagramm. Die Funktion, sein Ergebnis mit anderen Teilen zu können, soll einen weiteren Anreiz schaffen, ein gutes Ergebnis zu erreichen. Das Erreichen der Skulpturen ist durch die Gegebenheiten der Wege-Topologie ohne Umwege nur in gleicher Reihenfolge möglich, lediglich der Startpunkt der Wanderung spielt eine Rolle. Der Verlauf des Spiels gleicht somit einer Story mit immer neuen Events. Erfahrbar wird dies, neben den realen Eindrücken der Skulpturen, auch durch das Design der Gamification-Seite, die noch unerreichte Kunstwerke „verborgen“ aufführt. Ein zweites Gamification-Element und Träger eines vermittlungsbezogenen Prozesses ist die Anzeigeform der Fragen- und Anreize-Punkte auf der Landkarte rund um die Kunstwerke. Ungeöffnete Punkte sind in Rot, schon geöffnete Texte in Grün dargestellt. Die Farbänderung soll auf einfache, aber zielgerichtete Weise, Nutzer zur Auseinandersetzung mit allen Inhalten zu führen.

Das Design der Haupt-Gamification, das Sammeln von Badges durch den Besuch einer Skulptur, musste so verwirklicht werden, dass man sowohl selbst den „Punktestand“ beobachten als auch sein Ergebnis anderen klar ersichtlich machen kann. Um mehrere Ergebnisse anderer Besucher schnell überblicken zu können, wurde eine Präsentation gewählt, die schnell zu überschauen ist und möglichst ohne viel Text optisch den Grad des Erfolges vermittelt. Hierfür wurde zunächst eine Ansicht erstellt, die eine Art „Aktivierung“ einer Skulptur suggeriert. Sobald man sich in deren Nähe befindet füllt sich die Farbe der Skulptur und ein Anteil einer Art Tortendiagramm, dessen Felder farblich auf die Skulpturen abgestimmt sind. Wenn man sein Ergebnis teilen möchte öffnet man ein Formular in das persönliche Informationen eingetragen werden können. Der resultierende Eintrag im „Gästebuch“ setzt sich aus den eingegebenen Daten, dem automatisch ermitteltem Spiel-Ergebnis und dem Datum zusammen. Die Einträge werden in einer listenartigen Ansicht eingebettet, in der das Ergebnis-Diagramm große visuelle Bedeutung hat um einen schnellen Eindruck über alle Ergebnisse zu bekommen.

Die programmiertechnische Implementierung der aufgezeigten Inhalte sind mit dem nativen *Cocoa*-Framework für iOS, den Programmiersprachen Objective-C und Swift und dem Webservice *Google-Firebase* umgesetzt worden.

# 4 Evaluation

Zur Überprüfung der Funktionalität und der inhaltlichen (didaktischen) Substanz der WSW-App wurde ein Fragebogen zu den wesentlichen Aspekten der o.g. Komplexe erstellt. Für die Testphase wurde eine Gruppe von 7 Probanden mit iOS-Geräten und WSW-App auf dem WSW begleitet. Im Anschluss konnten die Probanden den Fragebogen zu den gemachten Erfahrungen ausfüllen. Die Befragungsergebnisse der Testgruppe (Tab. 1) lassen erkennen, dass die Teilnehmer sowohl bzgl. des Komplexes „allgemeine Benutzbarkeit“ als auch im Hinblick auf den Komplex zur „Auseinandersetzung mit dem Kunstprojekt“ bei der WSW-App einen schnellen Zugang gefunden haben. Danach ist die Navigation durch die App verständlich, selbsterklärend und bedienerfreundlich. Die Verknüpfung der implementierten Navigationsinhalte mit dem Audio-Guide sowie den Gamification-Elementen trägt einerseits zu einem interaktiven Umgang mit den jeweiligen Skulpturen bei und gibt andererseits, insbesondere durch das Spiel, einen Motivationsimpuls den ganzen WaldSkulpturenWeg ohne Orientierungsprobleme zu erwandern. Es deutet sich an, dass die Zielsetzungen der App als Erweiterung des museums-didaktischen Konzeptes des WSW erfüllt werden: Das Informationsportal zu den Skulpturen erschließt sich leicht, ist informativ und wirkt anregend, das Navigationsportal bietet Orientierung, das Gamification-System schafft Motivation.

| | | |
|---|---|---|
| Die Navigationsinhalte helfen, mich mehr auf die Skulpturen konzentrieren zu können. | trifft sehr zu:<br>trifft zu: | **6**<br>**1** |
| Die Hinweispunkte haben zu mehr Reflexion über die Skulpturen beigetragen. | trifft sehr zu: | **7** |
| Der Audio-Guide motiviert das Umherwandern im Skulpturumfeld. | trifft sehr zu:<br>trifft zu: | **4**<br>**3** |
| Das Spiel hat mich motiviert, den ganzen WSW zu erwandern. | trifft sehr zu:<br>trifft zu:<br>mittelmäßig: | **1**<br>**5**<br>**1** |
| Die Navigationsinhalte helfen, sich auf dem Weg zu orientieren. | trifft sehr zu: | **7** |

Tab. 1: Ergebnisse der Befragung.

Literaturverzeichnis

[CO14] Cook, M; van Riemsdijk, M.: Agents of memorialization: Gunter Demnig's Stolpersteine and the individual (re-)creation of a Holocaust landscape in Berlin. Journal of Historical Geography, Vol: 43, Page: 138-147, 2014.

[HE99] Hein, G. E.: Learning in the museum. Routledge, London, 1999.

[HE06] Hein, G. E.: A companion to museum studies. Blackwell, Chichester, 2006.

[HO91] Hooper-Greenhill, E.: Museum and gallery education. Leicester Univ. Press, 1991.

[KA14] Kapp, K.: Gamification: Separating Fact From Fiction. CLOMedia, 2014.

[MC17] McMunn-Tetangco, E.: Gamification: a practical guide for librarians. Lanham: Rowman & Littlefield, 2017.

[NI15] Nicholson, S.: A RECIPE for Meaningful Gamification. New York: Springer, 2015.

[RE95] Rese, B.: Didaktik im Museum: Systematisierung und Neubestimmung. Habelt, Bonn, 1995.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 259

# How Are You, Chatbot? Evaluating Chatbots in Educational Settings – Results of a Literature Review

Sebastian Hobert[1]

**Abstract:** Evaluation studies are essential for determining the utilization of technology-enhanced learning systems. Prior research often focuses on evaluating specific factors like the technology adoption or usability aspects. However, it needs to be questioned if evaluating only specific factors is appropriate in each case. The aim of this research paper is to outline which methods are suited for evaluating technology-enhanced learning systems in interdisciplinary research domains. Specifically, we focus our analysis on pedagogical conversational agents – i.e. learning systems that interact with learners using natural language. For instance, in addition to technology acceptance, further factors like learning success are more important in this case. Based on this assumption, we analyze the current state-of-the-art literature of pedagogical conversational agents to identify evaluation objectives, procedures and measuring instruments. Afterward, we use the results to propose a guideline for evaluations of pedagogical conversational agents.

**Keywords:** chatbots, pedagogical conversational agents, technology-enhanced learning, evaluation

## 1 Introduction

The rising of intelligent systems based on methods known from machine learning and artificial intelligence affects information systems in many areas. This is especially true for the domain of knowledge management and technology-enhanced learning. On the one hand, those technologies enable the analysis of massive amounts of data as it is used, for instance, for learning analytics [Ch12, Jü18]. Thus, artificial intelligence and machine learning algorithms improve the data analysis in this domain. On the other hand, the technologies enable the rise of new or massively improved systems for educational purposes. One topic that is especially promising and receives increasing interest are chatbot-based systems that use natural language processing algorithms [TP12]. In the domain of education, such systems are known as pedagogical conversational agents, i.e. chatbots used for educational purposes. Even though chatbots have been researched for many years, the application of machine learning and artificial intelligence algorithms enables new possibilities regarding natural language processing. One example in which major progress has been made due to machine learning is the so-called intend classification. The intent classification is an important part of the development of conversational agents as it is used to analyze natural language input with the aim of classifying it [Br17]. Due to these advances in natural language processing, the availability of easy-to-use libraries for chatbots increased rapidly. For this reason, researchers as well

[1] Universität Göttingen, Professur für Anwendungssysteme und E-Business, Platz der Göttinger Sieben 5, 37073 Göttingen, shobert@uni-goettingen.de

doi:10.18420/delfi2019_289

as practitioners were encouraged to develop conversational agents.

By analyzing the literature targeting pedagogical conversational agents, it is noticeable that researchers from different fields of research like computer science, information systems, education, and psychology are participating in this research stream. This interdisciplinarity results on the one hand in interesting complementary research results. On the other hand, there is no uniform procedure for evaluating the pedagogical conversational agent prototypes. The different research disciplines apply their own evaluation approaches to pedagogical conversational agents. Due to this, pedagogical conversational agents are often only evaluated regarding selected aspects. For instance, researchers in the information systems discipline often apply widely adopted models for technology acceptance or adoption like TAM [DBW89]. For analyzing usability in other fields, UEQ [LHS08] or meCUE [MR13] are common standardized questionnaires. In some cases, software prototypes are evaluated from a technical perspective whereas others focus on learning success. However, as the field of pedagogical conversational agents is interdisciplinary, it needs to be questioned if evaluating further aspects from other closely related fields might be beneficial for this interdisciplinary research domain. Thus, the aim of this research paper is to outline which commonly used evaluation models exist and which methods can be used for evaluating educational chatbots comprehensively. Based on this, we discuss possible implications for future research studies that aim at evaluating pedagogical conversational agents comprehensively targeting evaluation objectives from multiple perspectives.

As a starting point, we analyze the current state-of-the-art of evaluating pedagogical conversational agents based on a systematic literature review. Our focus is on identifying cases that have conducted evaluation studies. In particular, we are interested in the cases' objectives, procedures, and corresponding measuring instruments as well as discussing implications for future research:

*RQ1: Which state-of-the-art approaches to evaluate pedagogical conversational agents are common?*

*RQ2: Which implications can be derived for future evaluation studies of pedagogical conversational agents?*

To answer these two research questions, the remainder of this article is structured as follows: First, we briefly describe the theoretical background by defining the term pedagogical conversational agent and outlining related research. Based on this, we present the procedure of our systematic literature review that we applied to identify state-of-the-art evaluation approaches of the pedagogical conversational agent in the literature base. Following this, we discuss implications for future evaluation studies. Finally, we summarize our work in the conclusion.

## 2 Theoretical Background

Conversational agents – also known as chatbots, chatterbots, bots or interactive agents – that focus on educational purposes are known under different terms in this interdisciplinary research field. For instance, they are named pedagogical conversational agents [TP16]. These agents are a special form of chat applications that interact with learners automatically using a natural language interface [KHB07].

In recent research studies, especially messenger-like pedagogical conversational agents are popular [HM19]. This type of agents uses user interfaces known from instant messenger apps (like WhatsApp, Facebook or Telegram). In some cases, the pedagogical conversational agents are using APIs from those messenger platforms to integrate themselves into those platforms. In other cases, messenger-like agents are implemented as standalone apps and implement chat interfaces on their own. In contrast to messenger-like agents that are designed like messenger apps, pedagogical conversational agents can also be visualized using artificial bodies. This type of pedagogical conversational agents is called 'embodied conversational agent'. They often also interact with learners using text-based inputs. In some cases, also voice-enabled systems exist. The main difference to messenger-like agents is that the visualization does not only include a chat-interface but also shows an artificial body of the system. A common approach is to show a human-like avatar that is able to show facial expressions and gestures. According to [Pe16], messenger-like systems are currently preferred because "interaction takes place through messaging applications to which students are already very keen on" [Pe16].

From a technical perspective, pedagogical conversational agents are often similar to other implementations of chatbot-based systems. Usually, the user's natural language input is processed in the natural language understanding step. In this step, the user's input is often analyzed regarding known patterns in order to identify predefined intents. Based on the identified intents, information that is needed to respond to the user's input is processed before an answer is generated in the natural language generation step. Common approaches for handling natural language inputs are machine learning approaches, e.g., based on *spacy* or pattern matching approaches based on the *Artificial Intelligence Markup Language* (see e.g. [Mi09]).

In addition to conversational agents in other application domains, more advanced pedagogical conversational agents often store learning objects that encompass learning materials that can be requested by learners (see e.g. Co18). In some cases, even complex learning paths are stored that describe which learning objects should be learned next. To use this information, a pedagogical conversational agent needs to retain the learner's current state of knowledge to be able to deliver the next appropriate learning object. Due to this increasing complexity, these pedagogical conversational agents can also be classified as intelligent tutoring systems [CLW17].

## 3 Research Design

To answer our research questions, we conducted a literature search to identify research studies that evaluated pedagogical conversational agents. The resulting list of relevant studies creates the case base for the following analyses.

To create our case base, we conducted a structured literature search within major IS journals and IS-related databases and the digital library of the DeLFI proceedings. In particular, we searched the digital libraries of ACM, AIS, EBSCOhost, IEEE, and ScienceDirect as well as the AIS Senior Scholars' Basket of Journals [As18]. To search for literature, we used several different search terms to identify relevant papers using as outlined in Figure 1. Due to the focus of our study, we only included results targeting educational settings, which is reflected in the search terms. After conducting the literature search, we first checked titles and abstracts of search results and removed irrelevant articles. Then, we reviewed the full texts of the remaining articles for content fit using inclusion and exclusion criteria in accordance with [vo15]. Finally, we included all articles in our case base that describe evaluation studies.

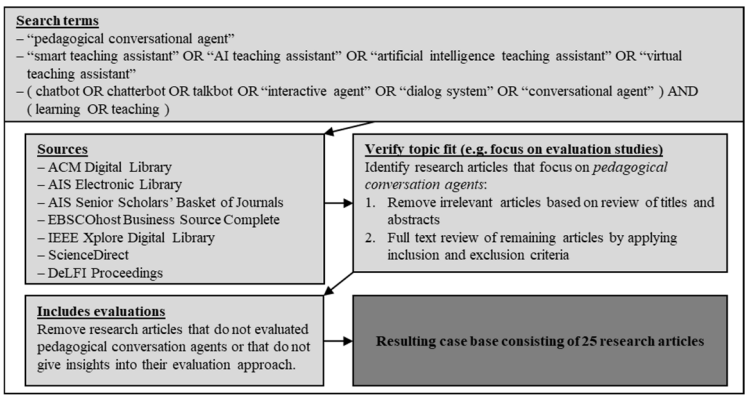



Fig. 1: Literature-based approach to create our case base

As a result, we got a case base consisting of 25 research papers that provides insights into the evaluation of pedagogical conversational agents.

## 4 State-of-the-Art Evaluation Approaches

We present the results of our analysis that outline how researchers evaluate pedagogical conversational agents. Before we describe the three dimensions "evaluation objectives", "procedures" and "measuring instruments" in detail, we give an overview of our classification matrix in the following section 4.1. A complete list of all considered papers including further information is available at https://publikationen.as.wiwi.uni-goettingen.de/getfile?DateiID=743.

### 4.1 Classification matrix

By applying the steps described above, we inductively identified seven characteristics of evaluation objectives, four characteristics of procedures and five characteristics of measuring instruments in our case base. Table 2 shows the result of our classification process. We categorized all 25 studies of our case base and highlight dominant intersections between procedures and evaluation objectives as well as measuring instruments and evaluation objectives in the matrix.

| | | Evaluation objective | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Acceptance & adoption | Learning success | Increased motivation | Further beneficial effects | Usability | Technical correctness | Further psychological factors |
| Procedures | Wizard-of-Oz experiment with a non-functional simulated pedagogical conversational agent | x | | | | x | | x |
| | Laboratory study with a functional pedagogical conversational agent under controlled conditions | x | x | x | x | x | | x |
| | Field study with a functional pedagogical conversational agent in the field | x | x | x | x | x | | x |
| | Technical validation of algorithms or software components without an experimental setup | | | | | | x | |
| Measurement instruments | Quantitative survey after a laboratory or field study with users of a pedagogical conversational agent (*Post*) | x | x | x | x | x | | x |
| | Quantitative survey before and after a laboratory or field study with users of a pedagogical conversational agent (*Pre/Post*) | x | x | x | x | x | | x |
| | Qualitative interview study with users of a pedagogical conversational agent | | x | | | x | | x |
| | Discourse analysis of transcripts of the interaction of learners with a pedagogical conversational agent | x | | | | | x | |
| | Analysis of technical log files produced during the usage of a pedagogical conversational agent | | | | | | x | |
| **Legend:** x indicates dominant intersections identified in the literature base | | | | | | | | |

Tab. 2: Classification matrix of our literature-based analysis that includes the dimensions procedures, measurement instruments and evaluation objective

### 4.2 Evaluation objectives

The objectives of evaluation studies influence the procedures to be applied and corresponding measuring instruments. In our analysis, we identified that many papers do not only focus on one specific evolution goal but often target multiple. In some cases, the authors of studies do not even state which objective they pursued. Due to these difficulties, we chose to allow the classification of one study in multiple characteristics. Thus, the total sum of evaluation objectives may be larger than the number of studies in our case base.

When the authors did not state the evaluation objective directly, we derived it manually from the result sections.

One of the most common evaluation objectives in our case analysis is to analyze the *technology acceptance and adoption*. Approx. 30 % of all analyzed studies focused on this main evaluation goal. In many cases, the analyses also include the evaluation of educational goals like *learning success*. Although analyzing the overall learning success seems to be the most important aspect of a technology-enhanced learning solution, only approx. 33 % of the studies focused on this evaluation goal. However, closely related evaluation objectives were analyzed in the case base: For instance, the effect of an *increased motivation* triggered by using a pedagogical conversational agent was analyzed. Researchers tested whether the interactive natural language-based conversation with a pedagogical conversational agent can result in an increased motivation [e.g. TP12]. Even though this is closely related to an increased learning success, we decided to emphasize this effect separately, because research studies do not necessarily try to measure whether the increased motivation also results in a higher learning success – also this assumption suggests itself. Further studies analyzed individual education-related goals that cannot be aggregated to a specific term. We opted to aggregate them under the generic term *further beneficial effects on learning processes* because the aggregated studies target individual but closely related positive effects of pedagogical conversational agents that are not explicitly attributable to an increased learning success or an increased motivation. An exemplary beneficial impact that has been analyzed is for instance that pedagogical conversational agents support learners to understand a given topic [e.g. KHB07].

In addition to these educational objectives, the *usability* of pedagogical conversational agents has been targeted in approx. 25 % of the studies. Although we would have expected that correlations between user-friendliness, acceptance and learning success exist, this was only analyzed rarely. In particular, interrelations between the technology acceptance, the actual system use, and the learning success would provide additional insights into the effectivity of providing pedagogical conversational agents to students. In this way, the results of TAM-based questionnaires could be verified with actual usage data of pedagogical conversational agents.

Another technical evaluation goal is to verify the *algorithmic or technical correctness* of software implementations of pedagogical conversational agents. Research items in this category validate for instance if the algorithmic implementation responds with a correct answer (see e.g. [PPR08]).

Finally, we identified evaluation studies that analyzed further *psychological factors* – in addition to motivation mentioned above – like the enjoyment (see [e.g. Fr17]) or the implication of the design of pedagogical conversational agents on the social interaction of participants with the agent.

Summarizing the evaluation objectives, the interdisciplinarity of the research stream becomes apparent in a large variety of objectives. Our classification of the evaluation objectives outlines that the goals targeted in research studies vary a lot from technical

validations of algorithms via design-oriented evaluations of user acceptance to psychological effects. Although these goals are very heterogeneous, they all have their justification and are therefore useful for a comprehensive examination of pedagogical conversational agents.

### 4.3 Procedures

While the total number of evaluation goals was quite large, we could aggregate the applied procedures in the case studies to only four different approaches:

First, we can distinguish evaluation studies using functional software implementations in comparison to early-stage non-functional software artefacts. A common approach for evaluating the concepts of pedagogical conversational agents at an early stage of development is to conduct so-called *Wizard-of-Oz experiments* [MHH12]. In these experimental setups, participants get access to a semi-functional prototypical implementation of a pedagogical conversational agent. The software often only consists of a mock-up-like user interface that can accept inputs of learners that are automatically redirected to a remote operator. In contrast to fully functional pedagogical conversational agents, complex functional implementations including natural language processing are usually still missing. Thus, the remote operator takes over all control tasks and responds to the users' input. A critical success factor of this evaluation approach is that the study participants are assuming that they are interacting with a fully operational pedagogical conversational agent. Using this evaluation approach, for instance, the learners' technology acceptance or the usability of a pedagogical conversational agent can be evaluated in an early stage of development (see [e.g. KHB07]).

In our case base, we identified two more types of studies that are common in other fields of research as well: *laboratory studies* and *field studies*. In contrast to Wizard-of-Oz experiments, researchers in our case base usually conduct both, laboratory and field studies, using fully-functional software implementations. However, as it was expected, in most cases the software implementations used in field studies are often more advanced than in laboratory studies. Thus, we conclude that the conduction of laboratory studies is suited to test prototypes in an early state of the development process. In contrast to that, pedagogical conversational agents can only be evaluated in a field study after the software is fully implemented and production-ready.

Besides these three procedures in which human participants interact with a software artefact, we identified *technical validations* of algorithms or software components that did not conduct any experiment with human participants. Evaluations in those research studies are often focused on verifying the correctness of algorithms. To do this, researchers, for instance, provide artificial natural language input to the system and examine whether it responds with a valid output.

Summarizing the identified evaluation procedures, we conclude that besides laboratory and field studies that are common in other domains as well, especially Wizard-of-Oz

experiments and technical validations are suited in research on pedagogical conversational agents. Particularly Wizard-of-Oz experiments seem important to assess the suitability of natural-language-based systems in an early stage of development. Algorithmic evaluations seem important as well, as the handling of natural language user inputs are currently a critical success factor for pedagogical conversational agents.

### 4.4 Measuring instruments

During analyzing the measuring instruments used in the cases of the pedagogical conversational agent literature, we could identify five different instruments.

The most prominent types of instruments are *quantitative surveys* before or/and after a field or laboratory study. In our classification matrix, we distinguish between surveys that ask the participants to complete a questionnaire after the study has finished (post-survey) and those who ask the participants to complete a questionnaire before and after it (pre- and post-survey). Whereas the post-surveys are often used to assess the technology acceptance or usability, pre- and post-surveys have the ability to evaluate the learning success by comparing the state of knowledge before the intervention with the state afterward.

As a counterpart to these quantitative instruments, researchers in two cases conducted *qualitative interviews* with participants of a study to get detailed feedback on the impact of using pedagogical conversational agents on the learning success and on resulting effects.

Finally, we identified two technical measuring instruments used in multiple cases: *Transcripts of dialogs* and *technical log files*. With a focus on evaluating the interaction among learners and pedagogical conversational agents, researchers in multiple studies stored all dialogues during evaluation studies. Afterward, the transcripts could be analyzed using discourse analyses to get insights into the technical correctness of the logic behind the pedagogical conversational agents as well as into the learners' behavior. In addition to that, researchers often collect technical log data in field experiments.

Summarizing the results from analyzing the measuring instruments, we conclude that most often quantitative data, as well as technical log data, are collected. Both can be analyzed statistically. In contrast to that, the analysis of transcripts using discourse analysis seems especially time-consuming as most researchers in our case base did it manually (e.g. by coding the answers of the conversational agent for correctness).

## 5 Discussion and Implications

The first research goal of our study (RQ1) was to analyze the literature base of pedagogical conversational agents to identify evaluation studies and to examine them. In our analysis, we aggregated the results to depict the evaluation objectives in seven categories, the

procedures in four categories and the measuring instruments in five categories. To derive implications for future research studies (RQ2), we carefully analyzed these categories and propose four evaluation steps based on the evaluation methods described above. By applying all four evaluation procedures, it is possible to cover all identified evaluation objectives. Thus, a comprehensive evaluation of pedagogical can be achieved.

(1) Wizard-of-Oz experiments are suited to gather early feedback of learners about a pedagogical conversational agent. The main advantage of this method is that it can be applied even before a functional software prototype has been developed – only a remotely controllable user interface is required. By conducting a Wizard-of-Oz experiment, the evaluation objectives *acceptance & adoption* and *usability* can be targeted. (2) Subsequently, an important step of developing pedagogical conversational agents is the technical implementation of the natural language understanding and subsequent answer generation. As this step is critical to the success of a pedagogical conversational agent, we follow the researchers in our case base who validated the technical algorithms in their studies. Consequently, we propose to conduct technical validations after the natural language processing component has been implemented. As a precondition, researchers must create a dataset consisting of valid input data (e.g. questions) and expected output data (e.g. in-tended answers). Using such a dataset, the algorithms can be tested for validity. To verify its performance, the transcripts of the dialogues (input data and resulting output data) as well as technical log files of the algorithms should be collected and analyzed. (3) Following the technical validation and the finalization of a first functional prototype of a pedagogical conversational agent, we propose to conduct a pre-test in a laboratory setting with the aim to test the agent with potential users. This gives researchers the possibility to get feedback concerning the learning processes, the related learning content as well as the suitability of the conversational dialog form from potential users before an agent is used productively on a large scale. Furthermore, laboratory experiments can be conducted even when the prototype is not yet fully functional and can only operate in a limited setting. In addition to the analyses of transcripts of dialogues and technical log files, post-experiment questionnaires should be handed out to the participants. In this way, users can express their opinions, for instance, targeting technology acceptance and adoption. We state that learning success should not be evaluated in this phase of the development of a pedagogical conversational agent. Analyzing the long-term learning success is usually a complex task that requires elaborated measuring instruments. (4) After a successfully completed laboratory experiment and the finalization of the software implementation, a field experiment gives researchers the opportunity to test the conversational agent in a real setting (like in [PP13]). The overall goal of this last evaluation step is to test the implications of using a pedagogical conversational agent towards the learning success. Due to this, it seems particularly important to us that the pedagogical conversational agent including the natural language processing is fully operational before conducting the field experiments. If a pedagogical conversational agent is not able to understand the intents of learners, they may refuse to use it directly. If, however, the agent is used in a large-scale field study, researchers can test whether the expected effects of pedagogical conversational agents (e.g. learning success, increased motivation or further psychological effects) occur. In

particular, the analysis of long-term effects – like learning success – must be measured in longitude studies with pre- and post-experiment surveys. Additionally, interviews with key users of the pedagogical conversational agent may give researchers the possibility to get more detailed feedback about the effects on the learning process. Finally, in accordance with the other evaluation steps, we also propose to save transcripts of dialogues and technical log files for an in-depth analysis of usage patterns and user engagement. The analysis of the transcripts of dialogues also gives researchers the possibility to assess the most important learning contents (i.e. the contents that are mentioned most often in the chat dialogues). Thus, it is possible to identify key learning concepts as well as contents that are formulated in a misleading way.

| | | Evaluation objective | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Acceptance & Adoption | Usability | Learning Success | Increased Motivation | Further beneficial effects | Psychological Factors | Technical Correctness |
| Procedure | Wizard-of-Oz experiment | x | x | | | | | |
| | Technical validation | | | | | | | x |
| | Laboratory experiment | x | x | | | | | x |
| | Field experiment | | | x | x | x | x | |

Tab. 3: Comparison of the evaluation methods with the evaluation objectives

By applying these four steps as part of evaluations of pedagogical conversational agents, the identified objectives can be addressed (see Tab. 3). Thereby, researchers might be able to contribute not only to one field of research but to multiple subfields in this interdisciplinary research domain of pedagogical conversational agents. However, targeting multiple objectives from different fields of research at the same time might not be easy: This might be particularly true if multiple models (like TAM [DBW89]) and standardized questionnaires (like UEQ [LHS08] or meCUE [MR13]) should be considered at the same time. For instance, [HST18] indicated that using and comparing different standardized questionnaires might be difficult. Thus, in future evaluation studies, it might be interesting to analyze this in more detail.

## 6 Conclusion

The aim of our research study was to analyze the current state-of-the-art of evaluating pedagogical conversational agents in research studies in order to derive implications for future evaluation studies. Our case-based analysis of the literature depicts that there is a large number of different evaluation approaches common in research on pedagogical conversational agents. Nevertheless, our analysis shows that often only selected aspects are being analyzed. We attribute this to the interdisciplinary nature of the research area.

This leads to the fact that researchers only analyze particular aspects, which are in the focus of their discipline. Due to this, comprehensive evaluations that analyze pedagogical conversational agents from different perspectives are usually missing. To provide a guideline for future research, we proposed four evaluation steps that might be used as a guideline for evaluation studies. However, future research is needed to test if evaluating multiple objectives in one research study is practicable and provides adequate contributions.

Limitations of our research approach are based on the research design of our literature review. After conducting our structured literature review, we selected our cases as described above. As with any literature review, the assessment and classification of the relevant articles may be subjective in some cases. However, we tried to formalize it as much as possible. Nevertheless, our results contribute to research and practice as they provide a guideline that can be applied in future research studies as well as in projects of practitioners. Both target groups can benefit from the analysis as it provides a guideline of how to analyze pedagogical conversational agents during development processes.

## Bibliography

[As18] Association for Information Sytems: Senior Scholars' Basket of Journals, https://aisnet.org/page/SeniorScholarBasket, accessed: 12/07/2018.

[Br17] Braun, D. et al.: Evaluating Natural Language Understanding Services for Conversational Question Answering Systems. In (Jokinen, K. et al. ed.): Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue. Association for Computational Linguistics, Stroudsburg, PA, pp. 174–185, 2017.

[Ch12] Chatti, M. A. et al.: A reference model for learning analytics. International Journal of Technology Enhanced Learning 5/6 /12, pp. 318–331, 2012.

[CLW17] Crockett, K.; Latham, A.; Whitton, N.: On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees. International Journal of Human-Computer Studies 97/17, pp. 98–115, 2017.

[Co18] Coronado, M. et al.: A cognitive assistant for learning java featuring social dialogue. International Journal of Human-Computer Studies 117/18, pp. 55–67, 2018.

[DBW89] Davis, F. D.; Bagozzi, R. P.; Warshaw, P. R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 8/89, pp. 982–1003, 1989.

[Fr17] Fryer, L. K. et al.: Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. Computers in Human Behavior 75/17, pp. 461–468, 2017.

[HM19] Hobert, S.; Meyer von Wolff, R.: Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. In (Ludwig, T.; Pipek, V. ed.): 14. Internationale Tagung Wirtschaftsinformatik (WI 2019) Tagungsband, pp. 301–314, 2019.

[Jü18] Jülicher, T.: Education 2.0: Learning Analytics, Educational Data Mining and Co. In (Hoeren, T.; Kolany-Raiser, B. ed.): Big Data in Context. Legal, Social and Technological Insights. Springer International Publishing, Cham, pp. 47–53, 2018.

[KHB07] Kerly, A.; Hall, P.; Bull, S.: Bringing chatbots into education: Towards natural language negotiation of open learner models. Knowledge-Based Systems 2/07, pp. 177–185, 2007.

[LHS08] Laugwitz, B.; Held, T.; Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. In (Holzinger, A. ed.): HCI and Usability for Education and Work, 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20-21, 2008. Proceedings. Springer, Berlin, i.a., pp. 63–76, 2008.

[MHH12] Martin, B.; Hanington, B.; Hanington, B. M.: Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions. Rockport Publishers, 2012.

[Mi09] Mikic, F. A. et al.: CHARLIE: An AIML-based chatterbot which works as an interface among INES and humans. In: 2009 EAEEIE Annual Conference, pp. 1–6, 2009.

[MR13] Minge, M.; Riedel, L.: meCUE - Ein modularer Fragebogen zur Erfassung des Nutzungserlebens. In (Boll, S., Maaß, S. & Malaka, R. ed.): Mensch & Computer 2013: Interaktive Vielfalt. Oldenbourg Verlag, München, pp. 89-98, 2013.

[Pe16] Pereira, J.: Leveraging chatbots to improve self-guided learning through conversational quizzes: In: Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality, pp. 911–918, 2016.

[PP13] Pérez-Marín, D.; Pascual-Nieto, I.: An exploratory study on how children interact with pedagogic conversational agents. Behaviour & Information Technology 9/13, pp. 955–964, 2013.

[PPR08] Pilato, G.; Pirrone, R.; Rizzo, R.: A KST-BASED SYSTEM FOR STUDENT TUTORING. Applied Artificial Intelligence 4/08, pp. 283–308, 2008.

[TP12] Tamayo, S.; Pérez-Marín, D.: An agent proposal for Reading Understanding: Applied to the resolution of maths problems. In: 2012 International Symposium on Computers in Education (SIIE). IEEE, Piscataway, NJ, pp. 1–4, 2012.

[TP16] Tamayo-Moreno, S.; Perez-Marin, D.: Adapting the design and the use methodology of a pedagogical conversational agent of secondary education to childhood education. In: 2016 International Symposium on Computers in Education (SIIE). IEEE, Piscataway, NJ, pp. 1–6, 2016.

[vo15] vom Brocke, J. et al.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the Association for Information Systems 1/15, pp. 205–224, 2015.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 271

# Leseflüssigkeitstraining mit Amazon Alexa

Sara Durski[1], Lisa Frey[2], Marguerite Schwarzbach[3], Yasmin Ashimi[4], Ute Massler[5] und Wolfgang Müller[6]

**Abstract:** In diesem Beitrag wird die Konzeption, Entwicklung und Evaluation einer auf Amazon Alexa basierenden Anwendung (*Skill*), zum Trainieren der Leseflüssigkeit vorgestellt. Der Skill übernimmt im Mehrsprachigen Lesetheater, einer Lautlesemethode zur Förderung der Leseflüssigkeit, verschiedene Rollen. Damit haben Kinder die Möglichkeit, auch außerhalb des Schulunterrichts das laute Lesen einer oder mehrerer Rollen eines Lesetheaters zu üben, indem Alexa die Lesepartnerin darstellt. Durch iteratives Vorgehen wurde die Entwicklung des Prototypens fortlaufend evaluiert und angepasst. Eine abschließende Evaluation gab deutliche Hinweise, dass sich trotz verschiedener technischer Beschränkungen der Skill als Co-Partnerin für ein Lesetheater eignen kann.

**Keywords:** Amazon Alexa, Lesetheater, Leseflüssigkeit, intelligente Sprachassistentin

## 1 Einführung

Die Digitalisierung und fortschreitende Entwicklung neuer technischer Innovationen bringt neue Möglichkeiten für den Einsatz von Medien in Lehr-Lernkontexten mit sich. Haushalte, in denen Kinder heute aufwachsen, sind zu 98% mit mobilen Endgeräten ausgestattet und 97% haben Internetzugang [FPR17]. So kommen auch Kinder und Jugendliche damit von klein auf in Kontakt. Auch intelligente Sprachassistenten finden sich vermehrt in den Familien wieder. Bereits 8,7 Mio. NutzerInnen in Deutschland verwenden diese [GBE18]. Eine der bekanntesten ist Amazon Alexa. Durch ihre Interaktion, personalisierte Programmierungsmöglichkeiten und lebendige Stimme hat Alexa das Potential, auch im Bildungskontext gezielt eingesetzt zu werden. In diesem Projekt wurde eruiert, inwieweit sie sich als Lesepartnerin zur Leseförderung mit Kindern einsetzen lässt.

## 2 Problemstellung, Ansatz und Zielsetzung

Nach wie vor zeigen Studien zur Lesekompetenz klare Defizite bei Primär- und Sekundarstufenlernenden [Oe16]. Eine verminderte Lesekompetenz ist problematisch, da sie das

[1] PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, durskisara@stud.ph-weingarten.de
[2] PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, lisa.frey@stud.ph-weingarten.de
[3] PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, marguerite.schwarzbach@stud.ph-weingarten.de
[4] PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, yasmin.ashimi@stud.ph-weingarten.de
[5] PH Weingarten, Fach Englisch, Kirchplatz 2, 88250 Weingarten, massler@ph-weingarten.de
[6] PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, mueller@md-phw.de

doi:10.18420/delfi2019_285

Lernen in Sachfächern beeinträchtigen und zu schlechteren Schulleistungen führen kann. Empirisch belegter essentieller Bestandteil einer guten Lesekompetenz ist die Leseflüssigkeit [Ga09]. Flüssig lesen zu können bedeutet, über basale Lesefertigkeiten zu verfügen, die es ermöglichen, auf Ebene der Buchstaben, Wörter, Sätze und Textabschnitte Wortbedeutungen sicher und schnell zuordnen zu können, Texte mit einer angemessenen Lesegeschwindigkeit (vor-)zulesen und dabei sinnorientiert intonieren zu können [Ni00]. Verfahren des wiederholenden Lautlesens wirken sich positiv auf Genauigkeit, Flüssigkeit und Verständnis des Lesens aus [Ni00]. Viele Lautlesemethoden bedürfen jedoch eines Lesepartners sowie der intensiven Übung. Nicht jedem ist es jedoch möglich, das Lautlesen zuhause zu üben, da viele Eltern aufgrund von Berufstätigkeit oder mangelnder Sprachkenntnisse nicht als Übungspartner fungieren können. In diesem Projekt wurden eine Anwendung für Alexa, ein sogenannter Skill, entwickelt. Dieser übernimmt die Lesepassagen, die nicht durch die SchülerInnen selbst gelesen werden können.

## 3 Stand der Technik

Im Fokus dieses Projekts steht die Entwicklung und Evaluation von digitalen Lernmedien für die Methode des Mehrsprachigen Lesetheaters (MELT). Das Lesetheater integriert das wiederholte Lautlesen. Daher wird davon ausgegangen, dass es die Leseflüssigkeit fördern kann. Im Lesetheater trainieren kooperierende Schülergruppen das Vorlesen mit Hilfe dialogischer Texte, wobei Sprecher- und Erzählerrollen in der Gruppe aufgeteilt sind. Am Ende werden die geübten Texte in gestaltender Form im Plenum aufgeführt [Ni06]. MELT erweitert das Lesetheater um einen zwei- oder mehrsprachigen Ansatz, der neben den Schul- und Fremdsprachen auch Migrationssprachen integriert [Ku17]. Lernende und Lehrende schätzen das Mehrsprachige Lesetheater als sehr motivierend und lehrreich ein [GKM17]. Die Intensität des Übens sollte durch Selbstlernphasen gesteigert werden. Der Einsatz geeigneter medialer Unterstützungsangebote könnte dies leisten. Das hier präsentierte Projekt stellt daher eine Erweiterung des Mehrsprachigen Lesetheaters dar.

Existierende Technologien von intelligenter Sprachassistenten werden eingesetzt und adaptiert. Die bekanntesten sind derzeit von Apple (Siri), Google (Google Assistant), Microsoft (Cortana), Samsung (Bixby) und Amazon (Alexa). Amazon Voice Service bietet dabei die Möglichkeit, als Drittanbieter sogenannte Skills zu entwickeln und den Assistent um Kommandos zu erweitert. Echo Show bietet mittels eines Touchscreens erweiterte Interaktionsformen an. Sprachsysteme, wie Amazon Polly weisen zwar eine bessere Qualität auf, jedoch ging es bei diesem Projekt um das Gesamtsystem Alexa.

## 4 Methodik: Design-Based Research

Ziel unseres Projekts ist es, aus technischer Sicht die Grenzen und Möglichkeiten von Alexa als elektronisches Lernmittel in Form einer Übungspartnerin für MELT zu ermitteln. Dabei wurde die technologische Machbarkeit und Akzeptanz fokussiert. Im

Zentrum stand folgende Forschungsfrage: Inwieweit eignet sich Amazon Alexa als Lesepartnerin zum Trainieren der Leseflüssigkeit? Im Rahmen des Projekts wurde ein Design-Based Research (DBR) Ansatz verfolgt, in dem erkenntnis- und anwendungsorientierte Forschung in einem kontinuierlichen Kreislauf von Gestaltung, Durchführung, Überprüfung und Re-Design vereint wird [De03]. Ergänzend wurden Methoden und Instrumente des User Experience (UX) Designs und aus dem Usability Engineering [Di10] eingesetzt. Dabei wurden NutzerInnen bereits früh in die Produktenwicklung integriert, indem mit ihnen bereits während der Umsetzung regelmäßig die Gebrauchstauglichkeit mittels Usability Tests überprüft wurde.

## 5 Konzeption

Die Zielgruppe der Lesefördermaßnahme MELT sind SchülerInnen (Ende Grundschule, frühe Jahre Sekundarstufe) mit Förderbedarf im Bereich der Leseflüssigkeit. Folgende Zielsetzungen wurden verfolgt: (1) Ergänzung des Lesetheaters in der Schule mit Selbstlerneinheiten zuhause; (2) Übung des Lesens gewählter Rollen in einem Dialog; (3) Übernahme der Rolle eines Dialogpartners durch Amazon Alexa.

Bei der Verknüpfung von MELT mit Amazon Alexa soll von Vorteil sein, dass sie als Co-Partnerin für das Lesen von sämtlichen fehlenden Parts eingesetzt werden kann. Hinzu kommen die korrekte Aussprache und Betonung von Wörtern und Sätzen, wodurch ein Sprachvorbild gegeben wird. Durch die Verwendung von Alexa Echo Show kann das Display zum Ablesen der Texte fungieren. Vorteilhalt ist auch, dass Alexa zu den gemachten Fehlern beim Vorlesen Feedback geben kann. Sie addiert in einem Score die gemachten Fehler und gibt beim Beenden des Skills Rückmeldung über die Anzahl der Fehler.

Bereits zu Beginn des Projekts wurden Personas und Szenarios erstellt, welche stets im Fokus standen. Das folgende Szenario verdeutlicht die anvisierte Nutzung: Felix kommt aus der Schule. Dort kam im Englisch-Unterricht MELT zum Einsatz. Auch zuhause ohne seine MitschülerInnen will er seine Theaterrolle üben. Dazu verwendet er Alexa Show. Auf ihr kann er den Text ablesen und sich auf seine Rolle konzentrieren, während Alexa die Rollen der MitschülerInnen übernimmt. Bei fehlerhaftem Lesen gibt Alexa Feedback.

## 6 Umsetzung

Es wurde zunächst ein Prototyp für die eigene Alexa-Anwendung (*Skill*) programmiert. Ein Skill umfasst sowohl den Code (in Form eines cloudbasierten Dienstes) als auch die im Developer Portal bereitgestellte Konfiguration. Erstellung, Bearbeitung, Programmierung und das Testen dieser Skills ist über Amazon Developer Services [Am19b] und mit der AWS Lambda Console [Am19] möglich. Skills bieten damit die Möglichkeit, eine Geschichte angepasst vorlesen zu lassen, beispielsweise indem verschiedene Stimmen und Sprachen eingefügt werden. Der entwickelte Prototyp umfasst aktuell zwei Lesetheater-

stücke aus zur Verfügung stehenden Skripten [Ma17] (u.a. Dialoge aus „Heidi“ und dem „Dschungelbuch“). Im Codebeispiel (Abb. 1) ist die Erstellung des Templates für den Alexa Show Display und die dazugehörige Sprachausgabe von Alexa dargestellt.

Es zeigten sich folgende Probleme und Hindernisse. Alle Befehle (*Intents*) sind gleich angelegt und gleichberechtigt. Das heißt, dass jederzeit auf alle Intents zugegriffen werden kann. Dies hat den Vorteil, dass zu jedem Zeitpunkt an jeder Stelle im Text eingestiegen werden kann. Allerdings hat es auch den Nachteil, dass wenn Alexa etwas falsch versteht, sie in und zwischen den Geschichten springt. Hinzu kommt, dass es in vielen Sprachen von Amazon erst zwei oder drei verschiedene Stimmen gibt [Am19a]. Folglich müssen mehrere Rollen mit den gleichen Stimmen belegt werden. Ein weiteres Problem stellt sich bei der Alex Show heraus, durch die Berührung ihrer Displays hört sie auf vorzulesen. Während des Vorlesens sind damit keine Interaktionen möglich. Insbesondere kann Text nicht gescrollt werden. Bei Änderung oder Neu-Implementation von Geschichten können Hindernisse auftreten, da diese aufwendig sind und Programmierkenntnisse erfordern.

```
'HeidiVorlesenLos': function () {
    if (supportsDisplay.call(this)) {
        const bodyTemplate1 = new Alexa.templateBuilders.BodyTemplate1Builder();
        var template = bodyTemplate1.setTitle("GameLet App")
                                    .setTextContent(makePlainText("Du kannst Heidi
                                    .build();
        speechOutput = "Es geht los mit Heidi. Welche Rolle willst du übernehmen?
        this.response.speak(speechOutput);
        this.response.renderTemplate(template);
        this.response.listen();
        this.emit(":responseReady");
        }
```



Abb. 1: Codeausschnitt vom Skill (links) und Usability Test mit der Alexa App (rechts)

## 7 Evaluation

Der entwickelte Prototyp wurde mit drei Personen, die der Zielgruppe entsprechen, getestet und evaluiert. Dabei stand die Funktionsfähigkeit und Usability einer möglichen technischen Umsetzung im Fokus. Daraus wurden technische Optimierungen abgeleitet. Der Datenschutzaspekt wird insofern beachtet, dass keine Daten der Schüler gespeichert werden und davon ausgegangen wird, dass die Kinder Alexa nur für Lernzwecke nutzen. Nach einer kurzen Einführung in MELT und den Skill wurden die Probanden gebeten, selbstständig das Lesetheater durchzuführen (s. Abb. 1). Nachdem die Probanden beide Geschichten getestet hatten, wurde mit ihnen ein halbstrukturiertes Interview durchgeführt. Bei diesem wurden Fragen bezüglich der Akzeptanz, des Spaßfaktors und der zukünftigen Nutzung gestellt.

In allen drei Tests wurde die intuitive Nutzung des Skills begrüßt. Bei der Auswertung wurden folgende fünf Aspekte identifiziert: **Reaktion von Alexa:** Probanden empfanden die Reaktionszeit von Alexa und die Wartezeit als zu lang. Die Reaktionsempfindlichkeit von Alexa auf den vorgelesenen Text führte nur selten zu Problemen. **Verständnis:** Nach einer kleinen Einführung konnte beim Verständnis und der Nutzung sowie bei der Orientierung ein intuitiv positives Verhalten der Probanden vermerkt werden. **Funktion von Alexa:** Immer wieder finden sich kleine Funktionsfehler von Alexa, z.B. schaltet sie sich

ohne Grund aus. Bei den Lesetheaterstücken hätten sich die Probanden einen größeren Leseanteil gewünscht. Die Erweiterung von Alexa durch Echo Show wurde als sehr positiv angenommen: “Den Text von Alexa Show abzulesen, ist viel cooler als von Papier.“ **Aussprache:** Der Einsatz der verschiedenen Sprachen im Skill wirkte auf alle Probanden verwirrend, da Alexa eine Rolle mit verschiedenen Stimmen und in unterschiedlichen Sprachen spricht. Bei der Aussprache ist Alexa durchaus ein gutes Sprachvorbild. Leider kann sie die Aussprache der Probanden nicht berücksichtigen und lässt falsch betonte Sätze, nicht richtig ausgesprochene Wörter und fehlende Worte einfach stehen und liest bei einigen gravierenden Fehlern einfach im Script weiter, wenn ausreichend Wörter erkannt werden. **Spaß und Akzeptanz:** Probanden waren während der Testdurchführung mit Begeisterung dabei. In den Interviews wurde auf die Frage, ob sie den Skill auch in Zukunft zum Lesen-Üben nutzen würden, folgende Aussagen festgehalten: „Bei einem zweisprachigen Drehbuch würde ich die App zum Englisch lernen nutzen", „Ja, wann kann man sich die App für Alexa runterladen?“, „In der App sollte es viel mehr Bücher geben“.

Alexa fällt als sprachliches Vorbild aufgrund guter Betonung und deutlicher Aussprache der Wörter positiv auf. Allerdings versteht sie die Probanden unterschiedlich gut und reagiert bei fehlerhafter Aussprache der Probanden nicht immer angemessen. Wenn Alexa den Probanden tatsächlich nicht versteht, kann sie leider keinen genauen Fehlergrund angeben. Dadurch sind Unsicherheiten bei den Probanden entstanden. Dieses Problem hängt von der zur Verfügung stehenden Alexa-Software ab. Da auf alle Befehle zu jeder Zeit zugegriffen werden kann, wird der Einstieg in die zwei Lesetheaterstücke erleichtert. Diese Option kann aber auch zu Verwirrungen und Orientierungslosigkeit führen, wenn Alexa den Probanden falsch versteht und deshalb zwischen den Geschichten springt. Der Gaming-Charakter durch den eingebauten Fehlerscore motiviert zum einen die Probanden, zum anderen entsteht aber ein gewisser Leistungsdruck. Das schnelle Verständnis und die intuitive, richtige Nutzung der Alexa führten zu einem positiven Gefühl. Probanden waren sich in den Interviews einig und würden den Skill auch in Zukunft nutzen.

## 8 Zusammenfassung und Ausblick

Im Rahmen eines DBR-Ansatzes wurde die Frage adressiert, in welchem Rahmen Amazon Alexa als Leseassistentin im Kontext von MELT eingesetzt werden kann. Dazu wurde ein Prototyp eines Alexa-Skills durch iterative Tests konzipiert, umgesetzt und evaluiert. Es konnten die technischen Potentiale und Grenzen von Alexa als Übungspartnerin für MELT ermittelt werden. Trotz verschiedener Einschränkungen sind die Ergebnisse bislang positiv zu bewerten. Eine detaillierte Evaluation, in der ethische Fragen, der Datenschutz und insbesondere die sprachdidaktische Perspektive im Fokus stehen, steht allerdings aus. Durch Alexa Show sind weitere funktionale Ergänzungen möglich, die noch zu evaluiert sind z.B. die Darstellung von Worterklärungen oder Übersetzungen.

## Literaturverzeichnis

[Am19] Amazon Web Services, Inc. oder Tochtergesellschaft: AWS Lambda Entwicklerhandbuch, https://aws.amazon.com/de/, Stand: 14.03.2019, 2006-2019.

[Am19a] Amazon.com, Inc. oder Tochtergesellschaft: Amazon Alexa: Alexa Skill Kit, https://developer.amazon.com/de/docs/custom-skills/speech-synthesis-markup-language-ssml-reference.html#break, Stand: 14.03.2019, 2010-2019.

[Am19b] Amazon.com, Inc. oder Tochtergesellschaft: Amazon Alexa: Developer Account, https://developer.amazon.com/de/alexa-, Stand: 14.03.2019, 2010-2019.

[De03] Design-Based Research Collective: Design-Based-Research: An Emerging Paradigm for Educational Inquiry. Educational Researcher, 32/1, S. 5-8, 2003.

[Di10] DIN EN ISO 9241-210: Ergonomie der Mensch-System-Interaktion - Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme. Beuth, Berlin, 2010.

[FPR17] Feierabend, S.; Plankenhorn, T.; Rathgeb, T.: KIM-Studie 2016. Kindheit, Internet, Medien. Basisstudie zum Medienumgang 6- bis 13-Jähriger in Deutschland. Medienpädagogischer Forschungsverbund Südwest, Stuttgart, 2017.

[Ga09] Grabe, W.: Reading in a second language. Moving from Theory to Practice. Cambridge University Press, New York, 2009.

[GBE18] Gentemann, L.; Meyer, Böhm, K.; Esser, R.: Zukunft der Consumer Technology - 2018 Marketingentwicklung, Trends, Mediennutzung, Technologien, Geschäftsmodelle. Bitkom e.V., Berlin, 2018.

[GKM17] Götz, K.; Kutzelmann, S.; Massler, U.: Mehrsprachiges Lesetheater: Ergebnisse zur Akzeptanz dieser neuen mehrsprachigen Lesefördermethode. In (Kutzelmann, S., et.al. Hrsg.): Mehrsprachiges Lesetheater. Handbuch zu Theorie und Praxis. Budrich, Leverkusen/Opladen, S. 42-56, 2017.

[Ku17] Kutzelmann, S .et.al.: Mehrsprachiges Lesetheater. Handbuch zu Theorie und Praxis. Budrich, Leverkusen/Opladen, 2017.

[Ma17] Massler, U.: MELT Mehrsprachiges Lesetheater Erasmus+ Projekt, http://melt-multilingual-readers-theatre.eu/leseskripte/, Stand: 14.03.2019, 2014-2017.

[Ni00] NICHD – Nat. Inst. of Child Health and Human Development (eds.): Report of the National Reading Panel. Teaching children to read – An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. U.S. Government Printing Office, Washington DC, 2000.

[Ni06] Nix, D.: Das Lesetheater: Integrative Leseförderung durch das szenische Vorlesen literarischer Texte. Praxis Deutsch 33/199, S. 23–29, 2006.

[Oe16] OECD: PISA 2015. Ergebnisse im Fokus. https://www.oecd.org/berlin/themen/pisa-studie/PISA_2015_Zusammenfassung.pdf, Stand: 16.03.2019.

# Lehrfilme einfach einfach machen – erfolgreiche Konzeptionierung von Peer-to-Peer Lehrfilmen

Anja Pfennig[1]

**Abstract:** Lehrfilme im Kontext der Hochschulausbildung setzen einen audiovisuellen Impuls, werden als probates Lernmedium eingesetzt und umfassen eine Vielzahl unterschiedlicher Lernmethoden. Weil Lehrfilme in der Grundlagenlehre Werkstofftechnik als begleitende Lernmittel eingesetzt werden, müssen sie inhaltlich einem auf die Kursinhalte zugeschnittenen hohen Qualitätsanspruch genügen. Wichtig bei der Erstellung der Lehrfilme ist die Beteiligung von Studierenden (Peer-to-Peer Ansatz). Ihre unmittelbare Nähe zum modernen Lernprozess und die Anforderungen an Lernbedürfnisse sind entscheidend für den späteren Lernerfolg. Der Lernerfolg wird auch durch die Bereitstellung unterschiedlicher Filmformate unterstützt. Im Folgenden werden aus persönlicher Erfahrung heraus niederschwellige, praktische Empfehlungen zur Umsetzung von „funktionierenden" Lehrfilmen (keine Produktbewerbungen) gegeben und eine erste Bewertung des Einsatzes von Lehrfilmen in der Grundlagenlehre der Studieneingangsphase Maschinenbau vorgenommen.

**Keywords:** making-of, Lehrfilm, Lehrfilmformate, peer-to-peer

## 1 Konzeptionierung von Lehrfilmen an der HTW Berlin

Im Allgemeinen bewerten Studierende Lehrfilme als niederschwelliges Lernmaterial [KK12] mit hohem Stellenwert bei der Beurteilung ihres persönlichen Lernfortschritts [Kon15]. Lehrfilme im Grundlagenfach Werkstofftechnik setzen in der Studieneingangsphase einen audiovisuellen Impuls und umfassen eine Vielzahl unterschiedlicher Lernmethoden. Lehrvideos, deren Inhalte und Implementierung in der Lehre mit dem Lernziel des Kurses übereinstimmen, werden weitestgehend als eine hervorragende Unterstützung der Lehre und des Lernfortschritts verstanden [HC18]. Da der zeitliche Aufwand zur Erstellung guter Lehrfilme sehr hoch ist, lohnt sich deren Produktion nur, wenn sie wiederverwertet werden können. Für die Produktion guter Lehrfilme werden weder professionelle Filmkenntnisse noch -ausstattung benötigt [Cr19]. Wichtig ist, dass Studierende den wissenschaftlichen Kontext verstehen [HC18] und der Inhalt des Lehrfilms in das gewünschte Lernresultat einfließt, gut vorbereitet ist und klar dargestellt wird. In dieser Ausarbeitung werden wir uns nur auf "student-generated video"-Lehrfilme beziehen, die im sogenannten Peer-to-Peer Ansatz (von Lernenden für Lernende) erstellt werden [WF88], [Pf18]. Studierende werden bei der Erstellung der Lehrfilme einbezogen, da sie die Unterrichtsinhalte oft anders gewichten, als der/die Lehrende und sowohl das Erarbeiten des Inhalts als auch

[1] HTW Berlin, Hochschule für Technik und Wirtschaft, Berlin, Fachbereich 2 Ingenieurwissenschaften Technik und Leben, Studiengang Maschinenbau, Wilhelminenhofstraße 75A, 12459 Berlin, anja.pfennig@htw-berlin.de

doi:10.18420/delfi2019_98

das Format Lehrfilm den Lerneffekt durch positive Emotionen erhöht [PR17]. Jeweils 4-6 Studierende (i.d.R. ohne Filmerfahrung) sind im Rahmen des Fachübergreifenden Projekts (5 ECTS = 150 hours workload per person) in 5 Arbeitsphasen (Randbedingungen, Meilensteine sind vorab geklärt) eingebunden: 1. Erstellen des Drehbuchs, 2. Beschaffung von Bildmaterial oder Zeichnung von Bildern, 3. Vertonung, Text, Sprechertext, 4. Dreh bzw. Filmproduktion, 4. Schnitt, Editing und Postproduktio und 5. Fertigstellung und Einreichung (upload youTube).

Angestrebtes Ziel ist die Produktion von 4-5 Minuten Film pro Studierenden. Die Texte werden von Studierenden erstellt und meist von einer Person eingesprochen; für Bildmaterial findet sich i.d.R. ein*e Studierende*r mit Zeichentalent. Somit werden Urheberrechte (Bild, Texte, Stimme) unbedingt gewahrt und die Rechte der Filme liegen bei den Studierenden, die ihr Einverständnis zur öffentlichen Verwendung geben. Am Dreh und Schnitt sind wiederum alle Studierenden beteiligt, die insgesamt eine deutliche Verbesserung ihres werkstofftechnischen Gesamtverständnisses und hohes Detailwissen im Kerngebiet ihres „Filmthemas“ zeigen. Eine Evaluierung des Lernfortschritts dieser Studierendengruppen ist derzeit aufgrund der kleinen Personenzahl noch nicht aussagekräftig.

## 2 Making-Of

Bei der Produktion hat die einfache Handhabung aller Werkzeuge oberste Priorität, um sowohl Kosten und Produktionsaufwand zu reduzieren. Da der Inhalt des Lehrfilms im Vordergrund steht, sind bei der Produktion sind Kreativität und Mut gefragt. Technik und Bildqualität sind zweitrangig. Eckpunkte wurden auf der HEAd`18 [PM18] und ICR-TEL2019 [PM19] vorgestellt und hier kurz als Checkliste zusammengefasst:

1. Peer-to-Peer Ansatz: Einbindung von Studierenden in die Erstellung, um somit unmittelbar von Lernerfahrung zu profitieren
2. 5 min Filmlänge sollten nicht überschritten werden
3. KISS (keep it simple, stupid): Produktion, technische Ausstattung und Sprache
4. Das Drehbuch sollte perfekt sein, bevor die Umsetzung beginnt
5. Das Filmformat sollte zum Inhalt passen [PFE19]
6. Es sollte immer die Bedeutung des gesamten Satzes graphisch umgesetzt werden und nicht nur ein Schlüsselwort (z.B. Knochenmaterial hat eine hohe Festigkeit: nicht der Knochen, sondern in der Festigkeit liegt die Bedeutung).
7. Ein Smartphone ist ausreichend. Wichtig ist die Stabilisierung der Kamera.
8. Gute Ausleuchtung (Baustrahler) unter Tageslichtausnutzung und Weißabgleich
9. Aufnahmeraum (kein Hall) ist wichtiger als Mikrophon

10. Aufnahme: Seitenverhältnis: 16:9, Auflösung: min. HD 720p, min. 25 Bilder/Sek., ausreichend hohe Bitrate: Mbit/s bei 720p und 8 Mbit/s bei 1080p.
11. Aufnahme von Alternativen (Schnittauswahl) und kleinerer Zeitfenster (sprints)
12. Aufbau von sets in jedem beliebigen Raum, –einmal aufgebaut- nicht verändern!
13. open-source software für Schnitttechnik und Nachbearbeitung, bspw. Shotcut oder OpenShot (i.d.R. sind nur gängige Funktionen notwendig: Übergänge, Audiobearbeitung, Textwerkzeuge, einfache Animation, Export in gängige Formate)

Im Gegensatz zu Khan-Stil-Videos mit i. d. Regel improvisierten Texten (https://www.youtube.com/user/KhanAcademyDeutsch) ist die Reihenfolge: *Wort vor Bild!* während der Produktion von Dokumentationen (hier: werkstofftechnikspezifischen Grundlagenfilmen in Form eines visualisierten Lehrbuchs) entscheidend. Sobald der Text einmal eingesprochen wurde, sind alle nachträglichen inhaltlichen Veränderungen sehr mühevoll, beeinträchtigen die Tonqualität negativ und senken die Motivation der Studierenden erheblich. Der Inhalt sollte allein durch die Tonspur verständlich sein, nur dann ist die Sprache im wissenschaftlichen Kontext so gewählt, dass auch Anfänger folgen können. Hier hilft eine klare und präzise Sprache, mit eher kürzeren Satzkonstruktionen und einfachen Worten. Fachbegriffe sollten immer erläutert und Abkürzungen vermieden werden. Da der Inhalt zu 100% korrekt sein muss, sollte das Drehbuch in iterativen Schritten sorgfältig korrigiert werden. Im Drehbuch wird der Film in einzelne Szenen gegliedert. Bildebene und Tonspur werden in Worten und Bildern genau beschrieben: Dauer einzelner Szenen, was sagt der Sprecher, was ist zu sehen, wie ist die Darstellung der graphischen Szene, werden Texte eingeblendet, etc.. Skizzen der späteren Bilder/Graphiken sind ausreichend.

URL Lehrfilmbeispiele Studieneingang, Einführungen in die Thematiken (Überblicke):

- Einführung in die Werkstofftechnik WT; Legetrick (7 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5zu5oBgrsbpySESNdJ3Cupx*
- Einführung in Kunststoffe: Stop-motion (3 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5wUlfwge0VTxKokobD_OOK7*

URL Lehrfilmbeispiele Studieneingang, werkstoffspezifisch (inhaltliche Lerngrundlage):

- Struktur und Aufbau, Gitterfehler: (5 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5wlO3gea5jLFhxgAr3IiOja*
- Werkstoffmechanik, Smith-Diagramm: ppt-Animation (1 Film) *https://www.youtube.com/watch?v=srONa_GTRQ8*
- σ-ε-Diagramm: ppt-Animation (3 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5zFekrtTdwTn9jnQI5NRbHI*
- Wärmebehandlung: ppt Animation (3 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5zAaB_frRDs9PeMBnBYnCIu*
- Festigkeitssteigerung: stop motion (4 Filme) *https://www.youtube.com/playlist?list=PLUOlZMSZYz5wm7m-ahbD8r4dCjDU498mV*

## 3 Kurzbewertung Lehrfilmformate und Lehrkonzept

Bis heute wurden 56 Filme verschiedener Techniken bzw. Formate (PowerPoint-Animation, Legetrick, Realfilm, Zeichentrick, Vortrag und Stop-Motion) realisiert. Bei der Auswahl des Formats sollten drei Randbedingungen beachtet werden:

1. Für das Lernziel der Lehreinheit geeignet (Legetrick eignet sich nicht für komplizierte wissenschaftliche Zusammenhänge)
2. Personelle Unterstützung und technische Ausstattung (z.B.: Graphiker*in oder Illustrator*in, Studiotechnik, Regieunterstützung, Kamera, Mikrophon, Licht)
3. Motivation und Ressourcen der studentischen Gruppe

In Umfragen bestätigen Studierende, dass das Lehrfilmformat nahezu keinen Einfluss auf den "joy of use" eines Filmes und ihren persönlichen Lernfortschritt hat (Abb. 1); vielmehr hat jede Studierendengruppe individuelle Vorlieben. In drei voneinander unabhängigen Kohorten zeigte sich, dass Formate bevorzugt werden, in denen Menschen vorkommen (Hand, Person). Studierende fühlen sich dadurch persönlich angesprochen, was von erheblicher Bedeutung für den Lernprozess ist und bei der Auswahl des Formats beachtet werden sollte. Im Sommersemester 2019 wurden zum ersten Mal Lightboard-Vorlesungssequenzen mit in das Lehrfilmportfolio aufgenommen. (Anmerkung: auch die Konstruktion, Aufbau und Inbetriebnahme wurde von Maschinenbaustudierenden im SS2018-WS18/19 realisiert). Diese wurden sofort als sehr gute Unterstützung des „inverted classroom" Lehrkonzepts aufgenommen. Denkbar ist hier, dass Studierende hier ebenfalls in den Erstellungsprozess einbezogen werden (z.B. zur Darstellung von Fachbegriffen)

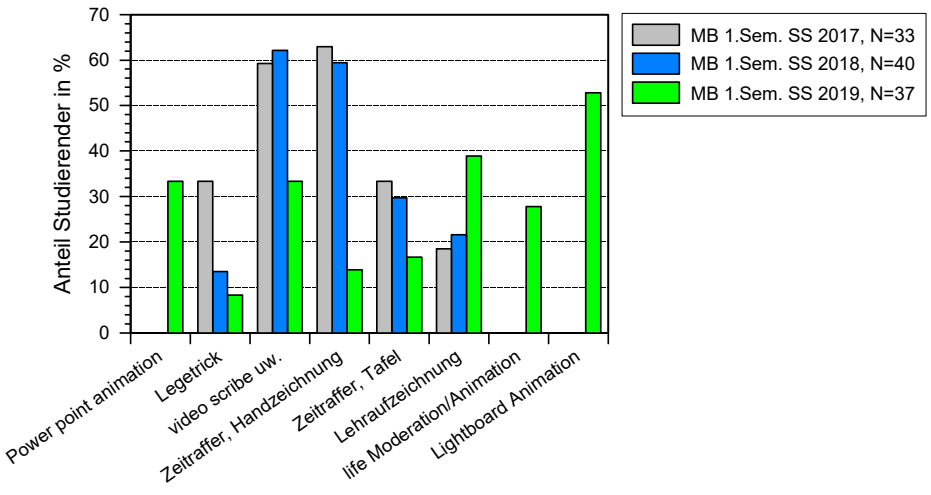



Abb. 1: Akzeptanz von Lehrfilmformaten

Vorteil der Lehrfilme sind Wiederholbarkeit, ständige Bereitstellung und „Portionierbarkeit“. Genau diese Vorteile rechtfertigen den sehr hohen Aufwand für die Produktion von Lehrfilmen, die in wiederholenden Grundlagenkursen immer wieder Anwendung finden, Lehrfilme dieser Art für ständig wechselnde Inhalte zu produzieren, ist nicht ratsam. Wichtig bei der Erstellung der Lehrfilme ist die Beteiligung von Studierenden (Peer-to-Peer Ansatz). Ihre unmittelbare Nähe zum modernen Lernprozess und die Anforderungen an Lernbedürfnisse sind entscheidend für den späteren Lernerfolg. Der Lernerfolg wird auch durch die Bereitstellung unterschiedlicher Filmformate unterstützt, die unterschiedliche Lernformen ansprechen und statt Monotonie die Vielfalt unterschiedlicher Lernmedien bieten. Die hohe Akzeptanz als Grundlage für „inverted classroom“-Lehrszenarien (Abb. 2 und Evaluationsberichte), die gute Vorbereitung der Studierenden und die Inklusion von Studierenden nicht-deutscher Herkunft mit Sprachschwierigkeiten in die Fachthematiken bestätigen diese Lehrfilme als vollwertiges Lernmaterial (auch wenn im SS2019 50% der Studierenden Präsenzunterricht bevorzugen). Mit großem Konsens werten Studierende Lehrfilme als gute aber zeitaufwändige Vorlesungsvorbereitung, deren inhaltlicher Anspruch auf den Unterricht abgestimmt ist (nicht zu leicht/schwierig). Wichtig ist allerdings, dass Lehrfilme unbedingt analog der Lehrinhalte produziert und angeboten werden. Lehrfilme, die wahllos empfohlen werden, lenken vom Unterrichtsschwerpunkt ab und sind eher ein Hindernis für einen gezielten Lernfortschritt.

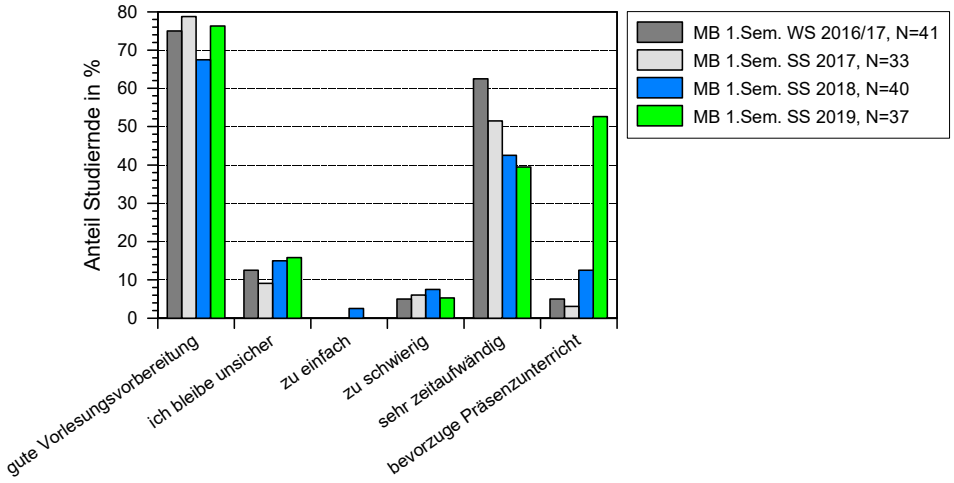



Abb. 2: Bewertung von Lehrfilmen und „inverted classroom“ Lehrszenarien

Die Heterogenität der Studierenden im ersten Semester mit wechselnden Anteil Studierender mit Sprachschwierigkeiten lässt eine Leistungsbewertung vor und nach Einführung von Lehrfilmen und „inverted classroom“ Lehrszenarien nicht zu. Obwohl sich die Noten über 13 Jahre nicht signifikant geändert haben, wertschätzen Kollegen höherer Fachsemester die Selbständigkeit der Studierenden sowie ihr Vermögen Zusammenhänge herzustellen und Transferaufgaben hinsichtlich werkstofftechnischer Grundlagen zu meistern.

## 4 Zusammenfassung

Lehrfilme können erfolgreich im Rahmen von angeleiteten Projekten von Gruppen von Studierenden produziert werden. Filme verschiedener Formate, wie PowerPoint-Animationen, Realfilm, Screencast, Legetrick, können so auf sehr hohem Niveau realisiert werden. Inhaltlich können alle Themen eines Kurses aufgegriffen werden und als Selbstlernmaterial produziert, bereitgestellt und in "inverted classroom" Lehrszenarien eingesetzt werden. Um das Interesse von Studierenden zu wecken und als Lerngrundlage akzeptiert zu werden, müssen Drehbücher und Filmprototypen sorgfältig korrigiert werden, da jedes Wort und jeder Satz korrekt sein müssen. Für die Produktion wird keine professionelle Ausstattung benötigt: nicht der perfekte Film, sondern das Lernziel des Kurses steht im Vordergrund. Studierende werten Lernfilme als vollständiges Lernmaterial und die verschiedenen Lehrfilmformate als lernförderlich, bewerten gleichzeitig Vorlesungsaufzeichnungen und Lightboard-Vorlesungssequenzen als besonders zuverlässig. Erfolgreiche Lernmedien sind Lehrfilme demnach nur, wenn diese das analog zum Lernziel des Kurses stehen und die Inhalte der jeweiligen Lerneinheit aufgreifen.

Literaturverzeichnis

[Cr19] Cresswell, S. L.; Loughlin, W. A.; Coster, M. J.; Green, D. M.: Development and Production of Interactive Videos for Teaching Chemical Techniques during Laboratory Sessions. Journal of Chemical Education 96/5, pp. 1033-1036, 2019.

[HC18] Hawley R.; Cate Allen, C.,: Student-generated video creation for assessment: can it transform assessment within Higher Education?. International Journal for Transformative Research 5/1, pp. 1-11, 2018.

[KK12] Kay, R.; Kletskin, I.,:Evaluating the use of problem-based video podcasts to teach mathematics in higher education. Computers & Education 59/12, pp. 619–627, 2012.

[Kon15] Kon, H.; Botelho, M.G.; Bridges, S.; Chiu Man Leung, K.: The impact of complete denture making instructional videos on self-directed learning of clinical skills. Journal of prosthodontic research 59/15, pp. 144–151, 2015.

[PM18] Pfennig, A.; Maier-Rothe, J.: Successfully planning and implementing Peer-to-Peer lecture films – Making-of. 4th International Conference on Higher Education Advances. Proceedings of the HEAd´18 1/18, pp. 45-54, 2018.

[PM19] Pfennig, A.; Maier-Rothe, J.: "Making of" – Successfully Planning and Implementing peer-to-peer Lecture Films. ICRTEL 2019 – International Conference on Research in Teaching, Education & Learning, 24-25 Feb., Dubai, 2019.

[PR17] Pirhonen, J.; Rasi P.: Student-generated instructional videos facilitate learning through positive emotions. Journal of Biological Education 51/3, pp. 215-227, 2017.

[WF88] Whitman, N.A.; Fife, J.D.: Peer Teaching: To Teach Is To Learn Twice. ASHE-ERIC Higher Education Report No. 4 ,1988.

# An E-Learning Tool for Parsing-as-Deduction

Anika Westburg[1]

**Abstract:** This paper presents the CL-Taskbox, an e-Learning tool for learning and teaching parsing-as-deduction, which is a specific approach to parsing. The tool provides parsing tasks the user solves with the help of constant visual feedback of the tool. The CL-Taskbox generates simple random tasks or allows to solve custom tasks, which the user can share and bookmark. The implementation is available online and anybody can use it with any common web browser.

**Keywords:** Computational Linguistics, e-Learning, context-free grammar, formal grammar, Parsing

## 1 Introduction

The CL-Taskbox is an e-Learning tool for parsing-as-deduction, which is a formalized approach to parsing with formal grammars in general [Pe83]. This approach formalizes parsing as a derivation. It defines the items that represent a step and the deduction rules which state how new items are generated from existing ones. It does not care about implementation details - the control structures are always the same, while the set of deduction rules defines the applied parsing algorithm. The tool makes the parsing-as-deduction approach observeable and thus allows for general understanding how parsing works and how the different algorithms relate to each other. The tool is interesting for students of computational linguistics and computer science when they learn about parsing, compiler construction or similar.

## 2 Features

The CL-Taskbox is a web-based implementation that anyone can access via a common web browser from every device with internet access without the need to install additional software. Under guidance of the tool the user solves parsing-as-deduction tasks that the tool randomly generates on-the-fly. For instance the user requests a random task for the Shift-Reduce algorithm for context-free grammars and the tool generates a grammar with production rules "N0 -> t1, N0 -> t0 N0, N0 -> t0, S1 -> N0" and start symbol "S1" to parse the input "t0 t1" with.

[1] Heinrich-Heine-Universität, Allgemeine Sprachwissenschaft, Universitätsstr. 1, 40225 Düsseldorf, anika.westburg@hhu.de

doi:10.18420/delfi2019_186

The exams of a parsing course might include similar tasks, hence the user can prepare for that as much and often as they want. At the moment the tool only generates simple grammars. On the other hand the user can insert a custom grammar and input and the URL reflects the task's data, so users can bookmark and share their favorite tasks or a lecturer can prepare tasks they want to show in a lesson or give their students to solve.

The user combines table items and production rules with the deduction rules to generate new table items. The tool highlights changing elements in green if the user can combine the element with the deduction rule or red if they cannot. Also the taskbox includes an automatic help system. The user can request that the taskbox uncovers the next step, solves the complete task or evaluates the result and says how many items out of all items the user generated. The tool Exorciser of ETH Zürich inspired the help system and the automatic task generation [Ts04].

Currently the CLTaskbox supports the following context-free parsing algorithms: Top-Down, Shift-Reduce, Left-Corner, Earley, CYK.

## 3 Outlook

The current implementation is a work in progress where the features named above have already been implemented, however, there is much more planned. The main feature to come is the generation of random tasks of several difficulty levels. As part of a dissertation the measurement of the difficulty of a task will be researched and the tool will collect usage data for analysis and evaluation.

More tasks for context-free grammars are to come: Left-Corner as Chart Version and Earley with passive items. Also tasks for tree-adjoining grammars and linear context-free rewriting systems will be implemented, concretely CYK and Earley for each. Also to make the tool widely available accessibility and mobile support are considered.

The tool is online hosted and accessible at [Da19].

Bibliography

[Da19] Daleh, S.: CL-Taskbox, http://samyadaleh.de, Stand: 13.06.2019

[Pe83] Pereira, F. C. N.; Warren, D. H.: Parsing as Deduction. In 21st Annual Meeting of the Association for Computational Linguistics, S. 137-144, 1983.

[Ts04] Tscherter, V.: Exorciser - automatic generation and interactive grading of structured exercises in the theory of computation, ETH Zurich, 2004.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 285

# Die intoMINT-App: Spielerisch für MINT begeistern und Berufsbilder vermitteln

Tobias Scheidat[1], Sandra Scholz[1], Mailyn Engel[1], Stephanie Böhnke[1], Heike Mrech[2] und Korinna Bade[1]

**Abstract:** Der deutschen Wirtschaft fehlen rund 338.000 Fachkräfte im MINT-Bereich. Gleichzeitig ist weiterhin eine mangelnde Chancengleichheit von Frauen und Männern beim Einstieg in zukunftsträchtige MINT-Berufe auf Grund stereotyper, gesellschaftlicher Einstellungen zu beobachten. Um Mädchen und Frauen zu motivieren, ihre berufliche Zukunft auf den MINT-Sektor auszurichten, entwickeln wir eine Smartphone-App, die spielerisch mittels Elemente der Gamification dazu anleitet, interessante MINT-Aktivitäten im realen Leben auszuprobieren. Diese Aktivitäten umfassen spannende Experimente und Projekte, die Beschäftigung mit unterhaltsamem Wissen oder den Besuch von interessanten MINT-Orten.

**Keywords:** MINT, Mädchen, Frauen, Gamification, Smartphone, App

## Die Smartphone-App intoMINT 4.0

Hauptziel des Projektes intoMINT 4.0 (www.intomint.de) ist die Entwicklung einer Smartphone-App, die dazu beiträgt, Schülerinnen ab der 8. Klasse für MINT-Themen zu begeistern und dies mit der Berufs- und Studienorientierung zu verknüpfen. Eine App als Medium wurde gewählt, um einen Großteil der Jugendlichen zu erreichen, da diese in der relevanten Altersgruppe (14 bis 18 Jahre) mit einem Anteil von ca. 90 % über ein eigenes Smartphone verfügen. Einen eigenen Computer besitzen dagegen nur ca. 70 % [Be17]. Die Grundidee der App ist eine geschlechts- und altersgerechte Verknüpfung von realer und digitaler Welt, indem den Mädchen in der App Anregungen gegeben werden, MINT-Aktivitäten im realen Leben umzusetzen. Dazu werden spannende Experimente vorgestellt, die zu Hause durchgeführt werden können, Wissen wird unterhaltsam und aktuell vermittelt und interessante MINT-Orte (bspw. Museum, offenes Labor) werden empfohlen. Durch die Möglichkeit, eigene Ergebnisse durch den Upload von Text und Bildern zu dokumentieren und zu präsentieren, werden reale und digitale Welt weiter miteinander verknüpft. Einerseits leiten erfahrungsbasierte Schwierigkeitsstufen durch die Aktivitäten der App, andererseits kann die Nutzerin verfügbare Aktivitäten nach ihren Vorstellungen und Kompetenzen selbst auswählen und durchführen. Die Verknüpfung mit MINT-Berufen, einer MINT-Ausbildung oder einem passenden MINT-Studiengang zeigt, wie vielfältig die Tätigkeitsfelder in diesem Bereich sind. Kern der App bilden Projekte

[1] Hochschule Anhalt, Fachbereich Informatik und Sprachen, Lohmannstr. 23, 06366 Köthen, vorname.name@hs-anhalt.de
[2] Hochschule Merseburg, Fachbereich Ingenieur- und Naturwissenschaften, Eberhard-Leipnitz-Straße 2, 06217 Merseburg, heike.mrech@hs-merseburg.de

doi:10.18420/delfi2019_279

aus dem MINT-Bereich, die mit wenig Aufwand und ohne bzw. mit geringen finanziellen Mitteln zu Hause durchgeführt werden können. Die Inhalte orientieren sich an der Lebenswelt der Mädchen, bspw. werden die MINT-Anteile bei Themen aus den Bereichen Gesundheit, Körperbewusstsein oder Umwelt hervorgehoben, um das Interesse der Mädchen zu wecken. Durch die selbstständige Bearbeitung der MINT-Themen werden die Mädchen spielerisch im wissenschaftlichen Arbeiten geschult. Dazu vermitteln wir eine positive Fehlerkultur, indem wir eigene Fehlversuche bei der Entwicklung der Experimente einbauen und im Wissenteil immer wieder Erfindungen vorstellen, die aufgrund eines Missgeschickes gelangen. So soll das Selbstkonzept der Mädchen gestärkt werden. Ebenfalls zeigen wir im Wissensteil, dass Interesse, Fleiß und Beharrlichkeit zu bahnbrechenden Erfolgen führen, da dies Eigenschaften sind, die sich die Mädchen selbst zuschreiben [Sc12]. Gamification Features vermitteln den Nutzerinnen, etwas geschafft zu haben und Fortschritte zu machen. Als Anerkennung erhalten sie z.B. Belohnungen in Form von Erfahrungspunkten, In-App-Objekten, Bewertungen oder Achievements. Ziel ist es, damit den Mädchen positive Erfahrungen im Umgang mit MINT zu ermöglichen und Anreize zu schaffen, sich weiter mit MINT zu beschäftigen. Weiter ist ein großes In-App-Event geplant, bei dem die in der App aktivsten Mädchen zu einer Veranstaltung eingeladen werden und reale Preise erhalten.

Im Projekt arbeiten wir mit einem lauffähigen Prototyp, der mittels agiler Software-Entwicklung schrittweise getestet (Stand Mai 2019: über 250 Tester*innen) und erweitert wird. Die App *intoMINT 4.0* wird für die zwei am häufigsten verwendeten mobilen Betriebssysteme in Deutschland, Android (ca. 71 % aller Smartphone-Nutzer) und iOS (ca. 28 %) [St19], mittels React Native entwickelt.

Das Verbundvorhaben „intoMINT 4.0“ der Hochschule Anhalt und der Otto-von-Guericke-Universität Magdeburg wird mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter den Förderkennzeichen 01FP1623 und 01FP1625 gefördert. Für den Inhalt der Publikation sind die Autoren verantwortlich. Ein wichtiger Kooperationspartner für das Projekt ist die Hochschule Merseburg.

Literaturverzeichnis

[Be17] Berg, A.: Kinder und Jugend in der digitalen Welt, 2017, url: https://www.bitkom.org/sites/default/files/pdf/Presse/Anhaenge-an-PIs/2017/05-Mai/170512-Bitkom-PK-Kinder-und-Jugend-2017.pdf, Stand: 31.05.2019.

[Sc12] Schmirl, J.; Pufke, E.; Schirner, S.; Stöger, H.: Das Zusammenspiel geschlechtsspezifischer Erwartungen, Einstellungen und Verhaltensweisen von Lehrkräften und Schülerinnen im MINT-Unterricht. In (Stöger, H.; Ziegler, A.; Heilemann, M., Hrsg.): Mädchen und Frauen in MINT. Bedingungen von Geschlechtsunterschieden und Interventionsmöglichkeiten. Lehr-Lern-Forschung. 1, Lit, Berlin u.a., S. 59–75, 2012.

[St19] Statista - Das Statistik-Portal: Marktanteile der führenden mobilen Betriebssysteme an der Internetnutzung mit Mobiltelefonen in Deutschland von Januar 2009 bis März 2019, url: https://de.statista.com/statistik/daten/ studie/184332/umfrage/marktanteil-der-mobilen-betriebssysteme-in-deutschland-seit-2009/, Stand: 31.05.2019.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 287

# Dungeons & DFAs

**Ein kollaboratives Multitouch Lernspiel zur Automatentheorie**

René Röpke[1], Matthias Ehlenz[1], Thiemo Leonhardt[2] und Ulrik Schroeder[1]

**Abstract:** In diesem Beitrag wird das Konzept und die Implementierung eines Multitouch Lernspiels zur Automatentheorie basierend auf dem an der RWTH Aachen entwickelten MTLG Framework vorgestellt. Es ist ein kollaboratives Lernspiel, welches unterstützt durch Digital Storytelling, einen spielerischen und zielgruppengerechten Einstieg in deterministische endliche Automaten als Themengebiet der Theoretischen Informatik gibt.

**Keywords:** Automatentheorie, Theoretische Informatik, Lernspiel, Multitouch, Tabletop

## Konzept und Implementierung

Themengebiete der theoretischen Informatik stellen für Lernende oft eine große Herausforderung dar. Die Komplexität von Beweisverfahren, die Abstraktheit formaler Notationen sowie fehlende mathematische Grundlagen mindern die Lernmotivation und auf lange Sicht das Interesse. Die Ergänzung klassischer Lehrmaterialien, wie Lehrbüchern und Arbeitsblättern, durch interaktive Elemente mit Fokus auf ein kollaboratives Lernerlebnis kann Motivationsfaktoren positiv beeinflussen und so einem Interessensverlust entgegenwirken.

Mit der Verbreitung von Multitouch-Displays sowohl in Bildungskontexten als auch in der alltäglichen Welt darüber hinaus, z. B. durch Smartphones und Tablets, kann der fehlenden Interaktivität in Lernprozessen innovativ entgegengewirkt werden. Da große Multitouch-Displays preiswerter und damit in Schulen interessant werden, können neben Einzelanwendungen auf kleineren Devices auch kollaborative Anwendungen für mehrere Spieler implementiert werden. Ein erster Lernspielprototyp zu Regulären Ausdrücken zeigt die Anwendbarkeit dieser Technologie zur Vermittlung von Inhalten der theoretischen Informatik [ERS18].

Das MTLG Framework ermöglicht die Implementation von Single- und Multitouch Lernanwendungen mittels moderner Webtechnologien [ELS17]. Es wird an der RWTH Aachen als Open Source Produkt entwickelt und in Softwarepraktika von Fach- und

[1] RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen, {roepke, ehlenz, schroeder}@informatik.rwth-aachen.de
[2] TU Dresden, Professur für Didaktik der Informatik, Nöthnitzer Str. 46, 01187 Dresden, thiemo.leonhardt@tu-dresden.de

doi:10.18420/delfi2019_320

Lehramtsstudierenden der Informatik zur Entwicklung von Lernspielen eingesetzt.

*Dungeons & DFAs* ist ein prototypisches Lernspiel für bis zu vier Lernende. Das Lernspiel bietet unterschiedliche Spielmodi zum kollaborativen Lernen von Inhalten aus der Automatentheorie. Als narrativer Anker (Anchored instruction) werden die Lernenden in eine Geschichte mit vier Helden versetzt, die im Auftrag des Königs eine Prinzessin aus einem Dungeon retten sollen. Der Weg im Dungeon erfolgt über ein Tunnelsystem, welches Höhlenstücke in Form eines nicht-deterministischen endlichen Automaten (NFAs) miteinander verbindet. Die einzelnen Höhlen repräsentieren dabei Level mit unterschiedlichen Aufgaben zu deterministischen endlichen Automaten (DFAs), z. B. Erzeugung eines Wortes, welches der Automat akzeptiert oder Entscheidung, welcher der Automaten ein bestimmtes Wort akzeptiert (siehe Abb. 1).

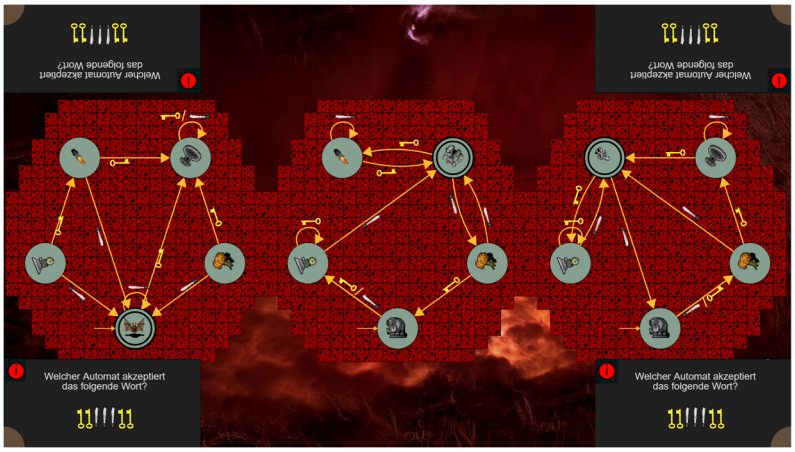



Abb. 1: Spielausschnitt aus Dungeons & DFAs

Kollaborative Interaktion wird in *Dungeons & DFAs* sowohl über das Eingabemedium, das Multitouch Displays, als auch über die unterschiedlichen Spielmodi gefördert. Die Lernenden müssen einstimmige Entscheidungen treffen oder gemeinsam ein akzeptierendes Wort für einen gegebenen Automaten eingeben.

Dieser Beitrag und die Demonstration zeigt die prototypische Umsetzung eines kollaborativen interaktiven Lernspiels für bis zu vier Lernende an einem Display.

Literaturverzeichnis

[ELS17] Ehlenz, M.; Leonhardt, T.; Schroeder, U.: Spielend leicht Lernspiele entwickeln - Ein Framework für Multitouch-Lernspiele. In (Igel, C. et al.): Bildungsräume 2017. Gesellschaft für Informatik, Bonn, S. 297-302, 2017.

[ERS18] Ehlenz, M.; Röpke, R.; Schroeder, U.: Kollaboratives Multitouch Lernspiel für Reguläre Ausdrücke. In (Krömker, D. und Schroeder, U.): DeLFI 2018 - Die 16. E-Learning Fachtagung Informatik, Gesellschaft für Informatik e.V., Bonn, S. 307-308, 2018.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 289

# E.V.A. – Emotionen Verstehen und Ausdrücken

Tobias Moebert[1] und Ulrike Lucke[1]

**Abstract:** Das Erkennen und Teilen fremder Emotionen sowie das Ausdrücken eigener Emotionen sind eine wesentliche Grundlage des sozialen Miteinanders. Diese Fähigkeiten sind bei Menschen mit Autismus beeinträchtigt. Das Training sozioemotionaler Kompetenzen ist deshalb gerade für Menschen mit Autismus nützlich. Ein solches Training kann mithilfe einer spielbasierten Anwendung effektiv gestaltet werden. Der Beitrag stellt ein solches Trainingssystem vor, welches die Spezifika der Zielgruppe adressiert.

**Keywords:** Computergestütztes Training, Digital Game Based Learning, Autismus, Adaptivität

## Adaptives Trainingssystem für soziale Kognition

Menschen mit einer Autismus-Spektrum-Störung (ASS) zeigen Defizite sowohl beim Erkennen von nonverbalen sozialen Signalen als auch beim Wahrnehmen, Verstehen und Bezeichnen von Emotionen. Diese Defizite führen zu deutlichen Schwierigkeiten sich in sozialen Situationen zurechtzufinden. Jedoch konnte auch gezeigt werden, dass sozioemotionale Kompetenzen bei Menschen mit ASS grundsätzlich trainierbar sind [GBC06].

Im interdisziplinären Verbundprojekt EMOTISK wurde ein adaptives Spielkonzept entwickelt, welches verschiedene Minispiele im Sinne des Digital Game-Based Learning verbindet. Als Ergebnis entstand die E.V.A.-App für Android Tablets [ZML17][Moe19], die sich durch folgende Merkmale auszeichnet:

- dynamische Generierung von über 3 Mio. Trainingsaufgaben, basierend auf empirischen Erkenntnissen hinsichtlich der Schwierigkeit des Erkennens und Differenzierens von Emotionen
- automatische Bewertung der Performance von Lernenden und der Schwierigkeit von Aufgaben anhand eines adaptierten Elo-Algorithmus [Elo08]
- reduzierte und klare Darstellung ohne ablenkende Designelemente, angepasst an die Bedürfnisse der Zielgruppe

Der entstandene Prototyp konnte bereits mit äußerst positivem Echo potentiellen Nutzern vorgestellt werden. Zudem wurde er in einer klinischen Studie erfolgreich hinsichtlich verschiedener Metriken untersucht (SUS, NASA-TLX, UEQ, meCUE).

[1] Universität Potsdam, Institut für Informatik & Computational Science, August-Bebel-Str. 89, 14482 Potsdam, vorname.nachname@uni-potsdam.de

doi:10.18420/delfi2019_324

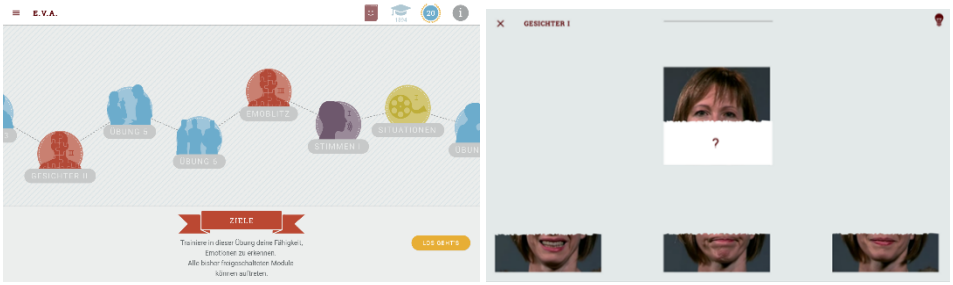



Abb. 1: Der Spielpfad (links) zeigt vergangene und zukünftige Spiel-Sessions an; bunte Sessions führen ein neues Minispiel ein. Jede Minispiel-Aufgabe (rechts) wird individuell generiert und ist an den Kenntnisstand des Spielers angepasst.

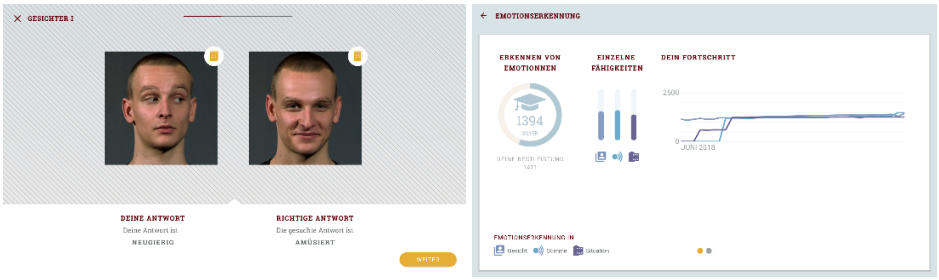



Abb. 2: Direktes Feedback erfolgt nach jeder Aufgabe (links). Der E.V.A.-Score (rechts) symbolisiert den Kenntnisstand des Spielers; er steigt oder fällt nach jeder Spiel-Session.

Das Projekt EMOTISK wurde vom BMBF unter Kennzeichen 16SV7241 gefördert. Wir danken Isabel Dziobek, Anne Weigand, Lioba Enk und Jan Schneider (HU Berlin) für die Mitwirkung beim Entwurf und beim Test der App.

Literaturverzeichnis

[Elo08] Elo, A. E.: The rating of chessplayers, past and present. Ishi, Mountain View, 2008.

[GBC06] Golan, O.: Baron-Cohen, S.: Systemizing empathy: Teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. In: Development and Psychopathology, 18:591–617, 2006.

[Moe19] Moebert, T.; Schneider, J. N.; Zoerner, D.; Lucke, U.: How to use Socio-Emotional Signals for Adaptive Training. In: Personalized Human-Computer Interaction. DeGruyter Oldenbourg, in print.

[ZML17] Zoerner, D.; Moebert, T.; Lucke, U.: IT-gestütztes Training sozio-emotionaler Kognition für Menschen mit Autismus. Informatik-Spektrum, 41(02):1556, 2017.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 291

# EduArc – eine verteilte Infrastruktur für OER

**Analyse zur Versionsverwaltung**

Nadine Schröder[1]

**Abstract:** Das Forschungsprojekt EduArc beschäftigt sich mit einer verteilten Infrastruktur, in der durch die Vernetzung von Repositorien die Auffindbarkeit und Nutzung von Open Educational Resources gefördert wird. In einem dezentralen System liegt ein Fokus auf der Handhabung von Versionen, die durch Nachnutzung und Bearbeitung von Materialien entstehen. Dieser Posterbeitrag beschreibt eine Analyse zur Versionsverwaltung.

**Keywords:** Open Educational Resources, EduArc, Infrastruktur, Repositorien, Versionierung, Versionsverwaltung

## 1 EduArc: Inhalte und Ziele

In dem vom Bundeministerium für Bildung und Forschung geförderten Projekt „EduArc“[2] entwickeln das Learning Lab der Universität Duisburg-Essen, das Leibniz-Institut für Bildungsforschung und Bildungsinformation (DIPF), das Leibniz-Informationszentrum Wirtschaft (ZBW) sowie die Universität Oldenburg gemeinsam ein erprobtes Gestaltungskonzept für eine verteilte Infrastruktur, um die Vernetzung und Verteilung von Open Educational Resources (OER) unter den deutschen Hochschulen zu fördern. Ziel dabei ist es technische, didaktische und organisatorische Anforderungen für eine vernetzte Bildungsarchitektur zu erforschen, indem unter anderem die Nutzung von Bildungsressourcen in der Lehre, OER-Infrastrukturen im Hochschulkontext sowie die Qualitätssicherung von OER untersucht werden.

Im Mittelpunkt des Projektvorhabens steht die Entwicklung einer verteilten Infrastruktur und deren Anbindung an verschiedene Lösungen und Systeme von Hochschulen. Durch die Vernetzung von Repositorien über ein zentrales System und den Austausch von Metadaten sollen Auffindbarkeit und Zugriff auf Ressourcen standortübergreifend ermöglicht werden. Der Core Hub verbindet sowohl institutionelle Repositorien einzelner Hochschulen als auch gemeinsam betriebene Repositorien, beispielsweise eines Bundeslandes. Ausgehend von den Repositorien sollen verschiedene Plattformen zur Erstellung und Ablage von Materialien angeschlossen werden können. [Ke19]

[1] Universität Duisburg-Essen, Learning Lab, Universitätsstr. 2, 45141 Essen, nadine.schroeder2@uni-due.de
[2] Laufzeit des Projektes: 01.10.2018-31.03.2022

doi:10.18420/delfi2019_261

## 2 Versionierung

Das Learning Lab der Universität Duisburg-Essen beschäftigt sich unter anderem mit dem Umgang von Versionen, die in der dezentralen Nutzung und Bearbeitung von Materialien entstehen können und deren Konsolidierung. Um die Grundgedanken der Nachnutzung, Veränderung und Weiterverbreitung von OER zu berücksichtigen, liegt in einem verteilten System ein Fokus auf der Darstellung und Handhabung von bearbeiteten Materialien. Dabei stellt sich die Frage, wie modifizierte Dokumente angelegt, Abweichungen angezeigt und zusammengeführt sowie darüber hinaus zu den Urhebern zurückgeführt werden können.

Um sich dieser skizzierten Problematik anzunähern, greift das Forschungsvorhaben auf die Praktiken von Versionierung im Bereich der Softwareentwicklung zurück, wo Systeme (z.B. Git, Mercurial) und Plattformen (z.B. GitHub, GitLab, Bitbucket) für das verteilte Versionsmanagement und das kollaborative Entwickeln von Code und Software verbreitet sind. In einem ersten Schritt wurden grundlegende Strukturen und Prozesse von Versionsverwaltung analysiert und deren Bedeutung sowie Übertragbarkeit für Lernressourcen untersucht. Hierbei konnte unter anderem festgestellt werden, dass sich das Kopieren eines Projektes („fork“) bei der Nachnutzung von Lernmaterialien als vorteilhaft erweist, während die Zusammenführung von zwei Projekten („merge“) beim Remix von OER Anwendung finden kann. Das Einfügen von Änderungen („commit“) und die Verwaltung einer Versionshistorie spielen beispielsweise beim kollaborativen Erstellen von Materialien eine wichtige Rolle. Weiterhin sind bei der Entstehung neuer Versionen verschiedene Materialarten und Dateiformate einzubeziehen, die einhergehen mit drei Granularitätsstufen von OER, die sich von einzelnen Ressourcen über Kollektionen von Dokumenten oder Lehreinheiten bis hin zu ganzen Kurse unterscheiden lassen [KH15].

Aufbauend auf diesen und weiteren Aspekten, die mit der Versionierung im Rahmen von Lernressourcen in Verbindung stehen, wurden Lerntools (z.B. H5P, LearningApps, Scratch, SlideWiki, u.a.) identifiziert und deren Funktionalitäten in Bezug auf Prozesse der Versionierung analysiert.

Auf Grundlage dieser Ergebnisse sind im weiteren Verlauf des Forschungsvorhabens Anforderungen für eine verteilte Lerninfrastruktur abzuleiten und prototypisch umzusetzen.

Literaturverzeichnis

[Ke19] Kerres, M. et al.: EduArc. Eine Infrastruktur zur hochschulübergreifenden Nachnutzung digitaler Lernmaterialien. In: Synergie. Fachmagazin für Digitalisierung in der Lehre #07, S. 66 – 69, 2019.

[KH15] Kerres, M.; Heinen, R.: Open informational ecosystems: The missing link for sharing resources for education. In: International Review of Research in Open and Distributed Learning, 16 (1), S. 24 – 3, 2015.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 293

# EDUSPACE – Social Media Lernplattform mit Schwerpunkt Inklusion

Ala Naoufal[1], Johannes Konert[2] und Hicham Naoufal[3]

**Abstract:** Im Rahmen einer Abschlussarbeit an der Beuth Hochschule für Technik wurde die Social Media Plattform EDUSPACE entwickelt, um im inklusiven Unterricht an deutschen Schulen eingesetzt zu werden. Der Schwerpunkt dieser Plattform ist das Bereitstellen und Beziehen von Lehr- und Lernmaterialien für den binnendifferenzierten Unterricht und die gezielte Förderung von leistungsschwächeren Schülerinnen und Schülern der Sekundarstufe I und II.

**Keywords:** differenziertes Lernen, eLearning, Inklusion, kooperatives Lernen

## 1 Motivation

Mit Voranschreiten der Digitalisierung an deutschen Schulen und dem Inklusionsgesetz, sollte der Einsatz einer digitalen Bildungslösung diskutiert werden, um den wachsenden Herausforderungen der inklusiven Bildung gerecht zu werden [Ke13]. Mit diesem Beitrag, sowie dem Poster vor Ort werden die Umsetzung einer solchen Bildungslösung und der aktuelle Entwicklungsstand präsentiert. Das System soll einerseits dazu beitragen, dass Lehrende, durch den gemeinsamen Austausch von differenziertem Material, den inklusiven Unterricht qualitativ mit verringertem Zeitaufwand gestalten. Und andererseits soll das selbstorganisierte Lernen insbesondere von leistungsschwächeren Schülerinnen und Schülern (SuS) durch Bereitstellen von geeignetem Material unterstützt werden. SuS sollen durch Vernetzung gemeinsam an Lösungsfindungen arbeiten und belohnt werden können.

## 2 Designkonzept und technische Spezifizierung

Materialien werden in EDUSPACE in Lehr- und Lernmaterialien unterschieden. Erstere sind solche, die ausschließlich von Lehrenden bereitgestellt werden und untereinander ausgetauscht werden können. Dies geschieht beim Festlegen von materialtypischen Eigenschaften wie Jahrgang, Fach, Schwierigkeit (Differenzierung) und Dateityp beim Upload-Prozess. Letztere sind Materialien, die sowohl von Lehrenden, als auch von SuS kategorisiert und bereitgestellt werden können. Diese werden den SuS beim Lernen mit EDUSPACE angeboten. Diese Unterscheidung führte zur Unterteilung der Plattform in

[1] Beuth Hochschule für Technik, Luxemburger Straße 10, 13353 Berlin, alanaoufal@gmail.com
[2] Beuth Hochschule für Technik, Luxemburger Straße 10, 13353 Berlin, jkonert@beuth-hochschule.de
[3] Beuth Hochschule für Technik, Luxemburger Straße 10, 13353 Berlin, hicham24.1@gmail.com

doi:10.18420/delfi2019_351

zwei Funktionsbereiche für die jeweiligen Zielgruppen durch Definition von Zugriffsrechten. In beiden sind die Funktionalitäten zum Suchen und Bereitstellen implementiert. Neben diesen Kernfunktionen und Benutzerkonten, soll der Lernerfolg der Schüler durch die Umsetzung eines Belohnungssystems, ähnlich wie bei Foren wie XDA Developers[4], begünstigt werden. Das Belohnen baut, neben dem Bereitstellen, auf die Funktion auf, Fragen und Antworten zu einem Material verfassen zu können. SuS können sich für hilfreiche Antworten auf ihre Fragen und für nützliches Lernmaterial beim Bereitstellenden bzw. Helfenden bedanken. Das Betätigen der Bedanken-Schaltfläche erhöht den Zähler des/der Empfängers/in im Benutzerkonto. Auf einer gesonderten Übersicht, werden SuS mit den höchsten Zählerständen anerkennend gelistet.

Die Plattform ist als Web-Anwendung umgesetzt, um den Einsatz über jedes internetfähige Gerät zu ermöglichen. Das Design der Anwendung orientiert sich an den Vorgaben des Material Designs und setzt gesättigte, unterschiedliche Farbtöne und Piktogramme ein, um Schaltflächen gerade für Inklusionsschüler verständlicher zu gestalten. Auf technischer Ebene ist die Anwendung in eine Server-Komponente auf Grundlage von Nodejs mit der Anbindung einer MongoDB Datenbank, sowie in eine Client-Komponente unterteilt. Dieses in React/Redux umgesetzte Frontend arbeitet autark und kommuniziert mit dem Backend über einen API-Client. Für die Entwicklung wurde der MERN-Stack[5] verwendet. Screenshots von EDUSPACE sind über einen Link[6] einsehbar.

## 3 Ausblick

Im nächsten Schritt soll der Aspekt der Dateiverwaltung und der übersichtlicheren Zuordnungen der Materialien nach thematischen Vorgaben aus den Rahmenlehrplänen weiter ausgebaut werden. Zudem soll die Erstellung von Lerngruppen implementiert werden, um insbesondere die differenzierte Hilfestellung und Rückmeldung für leistungsschwächere Schüler zu verbessern. Längerfristig ist geplant eine KI-gestützte Empfehlung von Lehr- und Lernmaterialien nach bestimmten Lernmustern zu realisieren.

Literaturverzeichnis

[Ke13] Kern, H. et.al.: Leitfaden zur Feststellung sonderpädagogischen Förderbedarfs an Berliner Schulen, Senatsverwaltung für Bildung, Jugend und Familie https://www.berlin.de/sen/bildung/schule/foerderung/sonderpaedagogische-foerderung/fachinfo/leitfaden_foerderbedarf-2017_nov15.pdf, Stand: 10.03.2019

[4] Link zu XDA Developers: https://www.xda-developers.com/, abgerufen am 10.03.2019
[5] Link zum MERN-Stack: http://mern.io/, abgerufen am 10.03.2019
[6] Abrufbar unter https://osf.io/y5m3p

# Eine „umgedrehte“ Video-Nachbesprechung einer Leistungskontrolle im E-Technik-Grundstudium

Mathias Magdowski[1]

**Abstract:** Für eine Grundlagenvorlesung zur Elektrotechnik wurde das Konzept des Flipped Classroom oder Inverted Classroom auf die Nachbesprechung einer Leistungskontrolle zur Prüfungszulassung übertragen. Für jede Aufgabe wurde ein kurzes Video aufgezeichnet, in dem kurz ein möglicher Lösungsweg und die typischen Fehler diskutiert werden. Eine solche Art der Nachbesprechung ist zeitnaher nach der Leistungskontrolle möglich, individueller anschaubar und damit besser für heterogene Studierendengruppen geeignet. Eine Herausforderung bei der Produktion ist, den Inhalt der Aufgabe nicht direkt zu verraten, so dass gleichzeitig kein frei verfügbarer Katalog von Klausuraufgaben im Internet steht.

**Keywords:** Flipped Classroom, Inverted Classroom, YouTube, Mobile Learning, Feedback, Open Educational Resources, Erklärvideo

## 1 Hintergrund und Idee

Die Grundidee des Inverted Classroom ist die Entschlackung und bessere Nutzung der kostbaren Präsenzzeit an der Hochschule durch die Verlagerung von Lerninhalten nach Hause. In diesem Sinne wurde die Nachbesprechung einer Zulassungsklausur aufgezeichnet und den Studierenden als Video zur Verfügung gestellt. Die zugehörige Lehrveranstaltung Grundlagen der Elektrotechnik an der Otto-von-Guericke-Universität Magdeburg (OVGU) ist eine zweisemestrige Pflichtveranstaltung im Grundstudium für mehrere Ingenieurstudiengänge und wird von ungefähr 150 Studierenden besucht. Die zugehörigen Übungen werden nach der HAITI-Methode [JU13] durchgeführt.

Wie in einer üblichen Klausurnachbesprechung sollten für jede Aufgabe die möglichen Ansätze genannt und diskutiert sowie die grobe Lösungsidee skizziert werden (Motto: „Wenn man weiß, wie es geht, ist es ganz einfach.“). Außerdem sollte auf typische Fehler und Missverständnisse hingewiesen werden. So bietet sich die Aufzeichnung der Aufgabennachbesprechung nicht nur als direkte Rückmeldung an den aktuellen TeilnehmerInnenkreis der Lehrveranstaltung, sondern auch als zusätzliche Ressource für zukünftige Studierende ähnlicher Lehrveranstaltungen an. Die Nutzung ist dabei nicht nur auf die OVGU beschränkt, sondern auch an anderen Hochschulen möglich.

[1] Otto-von-Guericke-Universität Magdeburg, Fakultät für Elektro- und Informationstechnik, Institut für Medizintechnik, Lehrstuhl für Elektromagnetische Verträglichkeit, Postfach 4120, 39016, Magdeburg, http://www.emv.ovgu.de/, mathias.magdowski@ovgu.de

doi:10.18420/delfi2019_286

## 2 Herausforderungen und Produktion

Die Herausforderung bei der Videoerstellung ist, die Antworten hinreichend genau zu erklären, ohne den Inhalt der Aufgaben direkt zu verraten, damit durch die Videos kein frei verfügbarer Aufgabenkatalog von Klausuraufgaben entsteht. Deshalb wurden einige Aufgaben geringfügig abgewandelt (z.B. gespiegeltes Diagramm, direkt umzeichnete und vereinfachte Schaltung, …), ohne den Lösungsweg zu verändern. Die Produktion der Videos erfolgte sehr schlank, wie bei dem unter [Lo19] beschriebenen Verfahren. Die Videos sind als Open Educational Resources unter CC-BY-Lizenz bei YouTube sowie auf einer Mediasite (universitätseigene Videoplattform der OVGU in Magdeburg) veröffentlicht. Sie werden außerdem im Lernmanagementsystem Moodle verlinkt und so den Studierenden angekündigt und zugänglich gemacht.

## 3 Vorteile und Auswertung

Eine solche Video-Nachbesprechung ist individueller, weil jede(r) Studierende nur die Videos schaut, die ihn oder sie interessieren, und somit besser für heterogene Studierendengruppen geeignet. Sie ist außerdem zeitnaher nach der Prüfung möglich, z.B. schon an darauffolgenden Tag und nicht erst beim nächsten Lehrveranstaltungstermin. Ein weiterer Vorteil ist die Möglichkeit besser vorbereiteter und detaillierterer Diskussionen bei der Prüfungseinsicht, da sich die Lehrperson die typische zehnminütige Erklärung bei jeder Aufgabe spart und direkt bei den „Knackpunkten“ einsteigt (Studierende: „Ich habe mir schon das Video angesehen, aber …“).

Interessant ist weiterhin, dass solche Videos von Studierenden nicht nur, wie ursprünglich gedacht, für die Nachbereitung von Leistungskontrollen sondern auch zur Vorbereitung auf summative Assessments genutzt werden. Denkbar ist auch, dass Studierende praktisch mittels Reverse Engineering anhand des beschriebenen Lösungswegs die zugehörige Aufgabenstellung entwickeln und anderen Studierenden zur Klausurvorbereitung zur Verfügung stellen. Obwohl man dadurch sehr viel über die Elektrotechnik lernen würde, ist dies nach Kenntnissen des Autors noch nicht vorgekommen.

Literaturverzeichnis

[JU13] Junige, M.; Umlauft, T.: „Und sie rechnen doch“ – Unterstützung der Selbstlernphase Studierender zur Vorbereitung von Rechenübungen. http://www.mathe.tu-freiberg.de/files/personal/111/junige-handout.pdf. Version: 03.10.2013

[Lo19] Loviscach, J.: Recording Technique. https://j3l7h.de/software.html. Version: 29.01.2019

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 297

# Einsatz von Chatbot-basierten Lernsystemen in der Hochschullehre - Einblicke in die Implementierung zweier Pedagogical Conversational Agents

Sebastian Hobert[1] und Florian Berens[2]

**Abstract:** Die universitäre Hochschullehre ist in vielen Bereichen geprägt von teilnehmerstarken, dozierenden-zentrierten Lehrveranstaltungen. Durch das oftmals unausgewogene Verhältnis von Studierenden zu Dozierenden bestehen Herausforderungen, Studierende in teilnehmerstarken Lehrveranstaltungen individuell zu unterstützen. Die vorliegende Demonstration gibt Einblicke in die technische Implementierung zweier Chatbot-basierter Lernsysteme, die diese Herausforderung adressieren, indem sie Studierende mittels Chatbots individuell unterstützen. Darüber hinaus können erste Erfahrungen aus deren Einsatz berichtet werden.

**Keywords:** Chatbot, Pedagogical Conversational Agent, Hochschullehre

## 1 Motivation

In der universitären Hochschullehre bestehen in der Studieneingangsphase häufig Herausforderungen, Studierende individuell zu betreuen und geeignete Hilfestellungen zu bieten. Dies ist besonders auf die häufig ausgeprägte Lehrform der teilnehmerstarken, dozierenden-zentrierten Lehrveranstaltung zurückzuführen. Durch vorlesungsbegleitende Zusatzangebote, wie bspw. Tutorien, versuchen Dozierende der fehlenden Interaktivität zu begegnen. Allerdings können diese Zusatzangebote den vielfältigen, individuellen Bedarfen der Studierenden oftmals nicht gerecht werden. Insbesondere in präsenzfreien Zeiten in der Vor- und Nachbereitungsphase von Präsenzveranstaltungen steht eine individuelle Unterstützung i.d.R. nicht zur Verfügung. In dem vorliegenden Forschungsansatz wird versucht, diesen Herausforderungen mittels Chatbot-basierter Lernsysteme (sog. *Pedagogical Conversational Agents; [TP16]*) zu begegnen. Die im Rahmen dieser Demonstration vorgestellten Chatbots verfolgen dabei das Ziel, Studierende in textbasierten Chats zu verstehen und individuelle Lernunterstützungsangebote bereitzustellen. Dazu werden zwei implementierte Chatbot-basierte Lernsysteme vorgestellt und im Rahmen der Demonstration erläutert, die (1) das Erlernen von Programmiersprachen unterstützen bzw. (2) zum Vermitteln von Statistik-Inhalten eingesetzt werden.

[1] Universität Göttingen, Professur für Anwendungssysteme und E-Business, Platz der Göttinger Sieben 5, 37073 Göttingen, shobert@uni-goettingen.de
[2] Universität Göttingen, Methodenzentrum Sozialwissenschaften, Goßlerstr. 19, 37073 Göttingen, florian.berens@uni-goettingen.de

doi:10.18420/delfi2019_288

## 2 Chatbot als Programmier-Tutor

Der erste vorgestellte Prototyp zielt auf eine verbesserte Lernunterstützung von Studierenden in Grundlagenveranstaltungen ab, in denen Programmiersprachen (bspw. in Java) vermittelt werden. Dazu wird ein ergänzendes Lernsystem bereitgestellt, dass eine webbasierte Programmierumgebung bereitstellt und Studierende darüber mittels eines Chatbots bei der Bearbeitung von Übungsaufgaben unterstützt. In Anlehnung an die Tätigkeiten menschlicher Tutorinnen und Tutoren, übernimmt der Chatbot die folgenden Aufgaben: (1) Textliche Erläuterung der Programmieraufgabe, (2) automatisches Beantwortung von Fragen zum Hintergrundwissens (z. B. zu Programmier-Konstrukten oder Algorithmen), (3) Bereitstellen von Schritt-für-Schritt Anleitungen, (4) Überprüfen des Quellcodes (automatisches Compilieren, Ausführen und Testen mittels JUNIT) und (5) proaktives Hinweisen auf Programmierfehler durch statische Quellcodeanalyse.

## 3 Chatbot als Unterstützung in Statistik-Vorlesungen

Der zweite Chatbot-basierte Prototyp dient der Unterstützung von Studierenden in teilnehmerstarken Statistik-Grundlagenvorlesungen. Dabei greift der als mobile Webapp implementierte Chatbot auf eine umfangreiche Datenbasis grundlegender Statistik-Fachbegriffe sowie Formeln zurück, um individuelle Fragen von Studierenden zu beantworten. Weiterhin ist der Chatbot in der Lage, den Studierenden passende Übungsaufgaben zu stellen, um das in der Vorlesung gelernt Wissen in präsenzfreien Zeiten zu vertiefen. Durch die Integration von Push-Notifications besitzt der Chatbot zusätzlich die Möglichkeit, proaktiv eine Kommunikation mit den Studierenden zu starten, um bspw. in Vorbereitung der nächsten Vorlesung wesentliche Kernaspekte zu wiederholen oder in Quizform abzufragen.

## 4 Ausblick

Für beide beschriebenen Anwendungsszenarien wurden prototypische, Chatbot-basierte Lernsysteme implementiert. Diese werden im Sommersemester 2019 mit Studierenden evaluiert, um detaillierte Einblicke in deren Wirkungsweisen zu erlangen, sodass bei der Demonstration der Prototypen erste Evaluationsergebnisse vorgestellt werden können.

Literaturverzeichnis

[TP16] Tamayo-Moreno, S.; Pérez-Marin, D.: Adapting the design and the use methodology of a pedagogical conversational agent of secondary education to childhood education. In: (García-Peñalvo, F.J.; Mendes, A.J. ed.): 2016 International Symposium on Computers in Education (SIIE), Salamanca 2016. IEEE, Piscataway, NJ, S. 1–6, 2016.

# Entwicklung eines E-Assessments basierend auf sequentiellem Testen für die sechsstufige Zensierung

Sima Darabi [1]

**Abstract:** In diesem Beitrag wird ein E-Assessment basierend auf sequentiellem Testen zur Ermittlung der Note in Rahmen eines sechsstufigen Notensystems vorgestellt. Die Anzahl der Fragen und die Prüfungsdauer steht nicht von Beginn an fest. Die Entscheidung, wann das Testverfahrens terminiert ist, hängt in jedem Schritt von der zuvor beantworteten Frage ab.

**Keywords:** E-Assessment, flexible Anzahl der Fragen, Flexible Zeit, Sequentielles Testen

## 1 Sequentielles Testen nach Wald

Beim herkömmlichen MC mit fester Anzahl von Fragen bestehen Alpha- und Beta-Fehler. Bei der Entstehung des Alpha-Fehlers wird ein unvorbereiteter Kandidat ungerechterweise als Könner betrachtet und bei einem Beta-Fehler ein Könner als Nicht-Könner betrachtet, der somit die Prüfung nicht besteht. Eine traditionelle Lösung für dieses Problem könnte die Erhöhung der Anzahl von Fragen sein [Sf78]. Wald (1963) stellt das Prüfverfahren mit sequentiellem Testen, das er 1947 als eine elegante Lösung für das Bestehen und Nicht-Bestehen einer Prüfung mit MC entwickelt hat, wie folgt vor: "Die sequentielle Analyse ist eine statistische Methode, deren charakteristisches Merkmal darin besteht, dass die Anzahl der Fragen nicht von Beginn an feststeht. Die Entscheidung für die Beendigung des Testverfahrens hängt in jedem Schritt an den zuvor beantworteten Fragen.“ [Wa63]

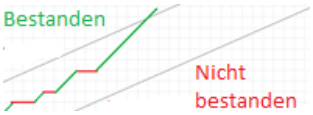



Abb. 1: Sequentielles Testen nach Wald

## 2 Entwicklung E-Assessment für mehrstufige Zensierung

Im Jahr 1975 haben Herbig und Erven (1975) das sequentielle Testen von Wald zu einer mehrstufigen Zensierung für die Wissensbeurteilung von Schülern weiterentwickelt [Hm76]. 1978 hat Sixtl (1978) in seinem Beitrag sequentielles Testen innerhalb der Pädagogischen Diagnostik in folgender Schlussbemerkung zusammengefasst: „Der

[1] SRH Berufsbildungswerk Neckargemünd, Im Spitzerfeld 25, 69151 Neckargemünd, sima.darabi@srh.de

doi:10.18420/delfi2019_66

sequentielle Wissenstest ist vergleichbaren Schulprüfungen herkömmlicher Art in mehrfacher Hinsicht überlegen“. [Sf78]

Das E-Assessments mit sequentiellem Prüfverfahren besteht aus einer Online-Fragenansicht, Online-Bewertungsansicht (Folgetestplan) mit einer Irrtumswahrscheinlichkeit von 0,1% sowie einem Pool gleichgewichteter MC-Fragen in der Datenbank zu einem Thema mit jeweils 4 Antwortmöglichkeiten, von denen eine richtig ist. Das Bewertungsansicht (Abb. 2, links) besitzt einen vordefinierte Rahmen mit 6 verschiedenen Notenbereichen (Noten 1 bis 6) [Hm76].

Die Schüler melden sich mit ihrem Account in der Fragenansicht an. Jeder Schüler bekommt eine zufällige Frage aus dem Pool. Beantwortet er die Frage richtig, wird auf der Bewertungsansicht gleichzeitig ein Strich senkrecht nach oben eingeblendet, und wenn er die Frage falsch beantwortet, wird ein Strich waagerecht nach rechts angezeichnet. Anfangspunkt ist die untere linke Seite. Die Schüler bekommen solange eine neue zufällige Frage aus der Datenbank, bis sie die Rahmen erreichen. Je nachdem, in welchem Bereich sie landen, erhalten sie eine entsprechende Note. Abb. 2 (rechts) illustriert den Verlauf von richtiger und falscher Antwort eines Teilnehmers, der nach 19 Fragen den zweiten Bereich verlassen hat. Damit ist die Prüfung zu Ende. Seine Note ist 2.

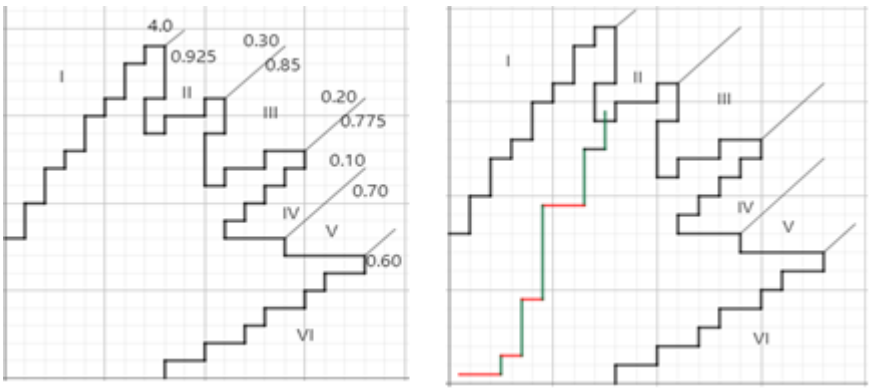



Abb. 2: Bewertungsansicht und Note 2 nach der Beantwortung von 19 Fragen

## Literaturverzeichnis

[Hm76] Herbig, M.: Studien zur Lehrforschung. Praxis lehrzielorientierter Tests. Pädagogischer Verlag Schwann, 1976; S.240

[Sf78] Sixtl, F.: Sequentielles Testen in der Pädagogischen Diagnostik. In (Klauer, K. J. Hrsg.): Handbuch der pädagogischen Diagnostik. Pädagogischer Verlag Schwann, Düsseldorf, 1978; S. 137–144.

[Wa63] Wald, A.: Sequential Analysis. Wiley, U.S.A, 1963; S.1

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 301

# Geocatch – Upsizing von geometrischen Konstruktionsaufgaben mithilfe einer Android-App

Eric Müller[1], Franz Schwarzer[2] und Karsten Weicker[3]

**Abstract:** Durch den Einsatz aktueller Smartphone-Technologie werden die für viele Schüler abstrakten mathematischen Inhalte der geometrischen Konstruktion zu einem Erlebnisspiel - verknüpft mit körperlicher Aktivität und Teamdynamik. Der Kurzbeitrag stellt das Konzept und einen ersten Prototypen vor und berichtet über die technischen Herausforderungen und Lösungen.

**Keywords:** Mathematik, Konstruktion geometrischer Figuren, mobile Anwendung, Gamification, GPS-Spiele.

## 1 Motivation und Konzept

Viele Schülerinnen und Schüler haben große Probleme mit den Inhalten des Mathematikunterrichts, da ihnen Verständnis und Anwendung abstrakter Konzepte und Vorgehensweisen schwer fallen. In diesem Paper wollen wir für geometrische Konstruktionsaufgaben einen Ansatz vorstellen, der sich die Tatsache zunutze macht, dass körperliche Aktivität die kognitiven Fähigkeiten positiv beeinflusst [BA11].

Konkret sollen Schülerteams die Konstruktion geometrischer Figuren in großem Maßstab nachstellen, indem sie beispielsweise ein Quadrat mit einer Kantenlänge von wenigstens 150m konstruieren, indem an jeder Ecke des Quadrats ein Schüler oder eine Schülerin platziert werden. Dies wird durch eine mobile App unterstützt und überprüft. In der App werden die Positionen der Teammitglieder in einer Übersicht angezeigt und als Hilfsmittel steht die Entfernungsmessung zwischen zwei Teammitgliedern sowie für einige Aufgaben auch die Winkelmessung zwischen drei Mitgliedern zur Verfügung.

Dieser Ansatz soll für Ausflüge oder Exkursionen geeignet sind und verknüpft die Beschäftigung mit praktischen mathematischen Fragestellungen mit einem Gaming-Konzept. Die grundsätzliche Machbarkeit von solchen Methoden haben Wijers et. al. [WJC10] mit ihrem MobileMath-Projekt gezeigt. Dort mussten von konkurrierenden Teams Parallelogramme in einer Fläche gestellt werden, wodurch Gebiete erobert werden und somit am Ende das Team mit der größten kontrollierten Fläche gewinnt. Das in unserem Paper vorgestellte Konzept, enthält deutlich weniger Gaming-Aspekte, versucht allerdings ein breiteres Aufgabenspektrum abzubilden.

[1] HTWK Leipzig, Fakultät Informatik und Medien, eric.mueller@stud.htwk-leipzig.de
[2] HTWK Leipzig, Fakultät Informatik und Medien, franz_ingo.schwarzer@stud.htwk-leipzig.de
[3] HTWK Leipzig, Fakultät Informatik und Medien, Institut für Informatik, Postfach 301166, 04152 Leipzig, karsten.weicker@htwk-leipzig.de

doi:10.18420/delfi2019_306

## 2 Aufgaben und Umsetzung

Einfache Einstiegsaufgaben sind gleichseitige oder gleichschenklige Dreiecke, Parallelogramme, Kreise mit ihrem Mittelpunkt oder auch Quadrate oder Rechtecke. Ist beispielsweise beim Quadrat nur die Entfernungsmessung erlaubt, müssen die Teammitglieder durch Messung der Diagonalen prüfen, ob es sich um rechte Winkel handelt.

Etwas schwieriger wäre beispielsweise die Konstruktion eines gleichseitigen Fünfecks nur mit der Entfernungsmessung – hier müsste über einen Kreis mit seinem Mittelpunkt konstruiert werden. Für Aufgaben wie die Konstruktion eines Kreises (durch drei Personen auf dem Kreis) und einer Tangente an den Kreis (zwei weitere Personen) wird die Winkelmessung benötigt. Auch das Aufstellen eines nicht gleichseitigen und nicht gleichschenkligen Dreiecks mit seinem Höhenschnittpunkt kann wesentlich einfacher konstruiert werden, wenn der benötigte rechte Winkel durch Winkelmessung geprüft werden kann.

Geocatch ist als Applikations-Stack bestehend aus REST-API, Android-App, Datenbanken und socketbasierter Kommunikation zwischen Client und Server umgesetzt. Es findet ein permanenter Austausch der Koordinaten der Spieler zwischen mobilen Geräten und Server statt. Die Ungenauigkeit der GPS-Koordinaten (vgl. [vdE15]) muss durch Mittelwerte bzw. Kalman-Filter ausgeglichen werden, um “Sprünge” der Spieler zu vermeiden. Über ein Regelwerk sind die zu akzeptierenden Lösungen beschrieben und die Positionen eines Teams werden diesbezüglich überprüft.

## 3 Fazit

Der Prototyp eines studentischen Softwareprojekt wurde im Rahmen dieses Papers weiterentwickelt und die grundsätzliche Machbarkeit gezeigt. Zukünftig sollen die Aufgaben erweitert werden und die App in ersten Versuchsklassen zum Einsatz kommen. Der Praxistest muss zeigen, ob ein Mechanismus zum manuellen Nachjustieren der eigenen Position im Rahmen der GPS-Ungenauigkeit notwendig ist.

Literaturverzeichnis

[BA11] Biddle, S.J.H.; Asare, M.: Physical activity and mental health in children and adolescents: A review of reviews. British Journal of Sports Medicine, 45, pp. 886-895, 2011.

[vDE15] van Diggelen, F.; Enge, P.: The World’s first GPS MOOC and Worldwide Laboratory using Smartphones. In: *Proc. of the 28th Int. Technical Meeting of The Satellite Division of the Institute of Navigation*, pp. 361-369, 2015.

[WJD10] Wijers, M.; Jonker, V.; Drijvers, P.: MobileMath: exploring mathematics outside the classroom. ZDM Mathematics Education, 42, pp. 789-799, 2010.

# Greifbares Lernen von Informatikinhalten

## Aktive Tangibles auf kapazitiven Multitouch-Tischen

Thiemo Leonhardt[1], Matthias Ehlenz[2]

**Abstract:** In diesem Beitrag werden zwei kollaborative Lernanwendungen unter Einbindung von Tangibles auf großen Multitouch-Displays im Bereich Informatik vorgestellt. Das erste Spiel behandelt Pattern Matching mit regulären Ausdrücken. Die Tangibles werden hierbei als Eingabegeräte genutzt und bieten zudem individuelle Feedbackmöglichkeiten. Das zweite Spiel behandelt das Erstellen und Simulieren logischer Schaltungen aus der praktischen Informatikausbildung. Hier ist besonders die Möglichkeit des kollaborativen Simulierens von Schaltungen über die Tangibles als logische Gatter ein wesentlicher Mehrwert.

**Keywords:** Tangibles, kollaborative Lernanwendung, Tabletop

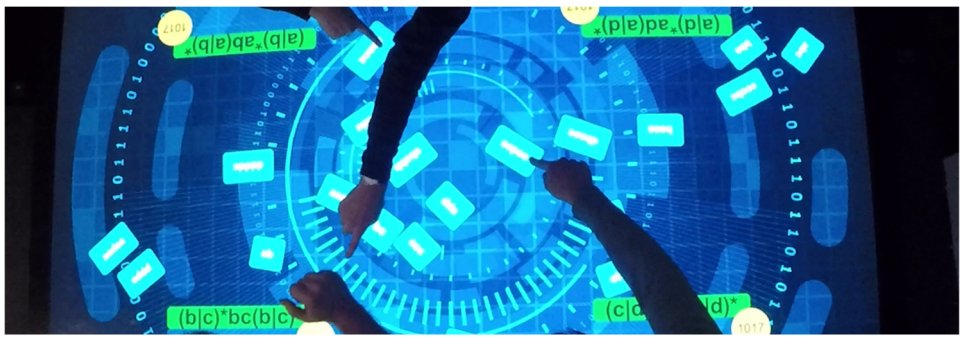



Abb. 1: Pattern Matching kollaborativ

## 1 Konzept

Die Förderung von Kollaboration bei der Erarbeitung von Unterrichtsinhalten wird im Sinne der Digitalisierung von Schulunterricht ein zentrales Thema. Die Verabschiedung vom klassischen Frontalunterricht sollte nicht zu einer Entwicklung von Einzelunterricht zwischen Lernenden und Tablet führen, sondern die kollaborativen Möglichkeiten elektronischer Lernhilfen in den Vordergrund stellen.

Das Forschungsprojekt TABULA ([Eh18], [Ch19]) soll das Erlernen von Informatikkonzepten greifbar und den Lernprozess damit erfolgreicher machen. Dazu

[1] Professur für Didaktik der Informatik, Technische Universität Dresden
[2] The Learning Technologies Research Group, RWTH Aachen University, Germany

doi:10.18420/delfi2019_347

werden große Multitouch-Tische (84“) mit kapazitiver Touch-erkennung von bis zu 100 Touch-berührungen gleichzeitig eingesetzt. Zusätzlich wurden für den greifbaren Part der Lernanwendungen sogenannte Tangibles entwickelt. Tangibles sind greifbare beliebige Eingabe- sowie Ausgabe-Objekte, die Bestandteil der Anwendung (zum Beispiel als Spielfigur eines digitalen Spiels) oder auch der Eingabe (zum Beispiel als Mausersatz, mit neuen Feedbackmöglichkeiten wie Vibration, Leuchten, Soundausgabe) sind. Die Konzentration auf große kapazitive Bildschirme ermöglicht es den Lernenden in einer Gruppe um den Tisch herum an konkreten Gegenständen Informatik-Konzepte gemeinsam zu arbeiten.

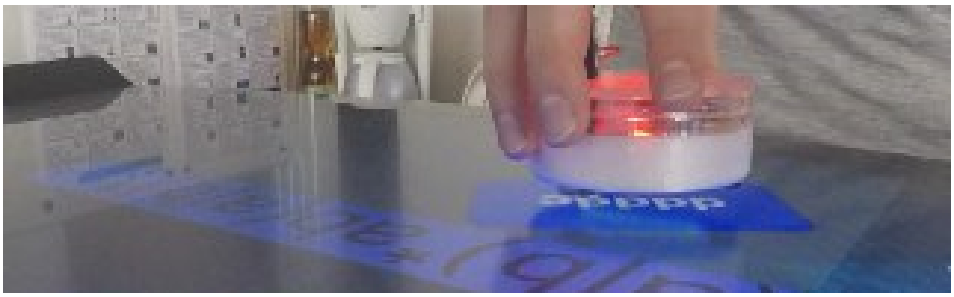


Abb. 1: Tangibles als Eingabe und Feedbackkomponente

In der Demonstration werden zwei kollaborative Lernanwendungen im Bereich Informatik vorgestellt. Das erste Spiel behandelt Pattern Matching mit regulären Ausdrücken. Der didaktische Aufbau ist jedoch auch auf Bereiche wie Chemie, Germanistik oder in Grundschulkontexte übertragen worden. Das zweite Spiel behandelt das Erstellen und Simulieren logischer Schaltungen. Hier ist besonders die Möglichkeit des kollaborativen Simulierens von Schaltungen herauszuheben. Beide Lernanwendungen nutzen die Tangibles einmal als Eingabe und Feedbackmöglichkeit und einmal als Teil der Simulation. Zielgruppe sind Studierende in der praxisorientierten Informatik-Hochschulausbildung, Auszubildende zum Mathematisch-Technischen-Assistenten sowie Schülerinnen und Schüler der gymnasialen Oberstufe im Fach Informatik.

## Literaturverzeichnis

[Ch19] Cherek, C. et al.: Off-Surface Tangibles: Exploring the Design Space of Midair Tangible Interaction. In (Dey, A.; Zhao, S. Hrsg.): Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, 2019.

[Eh18] Ehlenz, M. et al.: The Lone Wolf Dies, the Pack Survives?: Analyzing a Computer Science Learning Application on a Multitouch-Tabletop. In (Joy, M.; Ihantola, P. Hrsg.): Proceedings, 18th Koli Calling Conference on Computing Education Research. Koli Calling 2018 November 22-25, 2018, Koli, Finland. The Association for Computing Machinery, New York, 2018; S. 41–48.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 305

# HandLeVR - Handlungsorientiertes Lernen in der VR-Lackierwerkstatt

Raphael Zender[1], Matthias Weise[1] und Ulrike Lucke[1]

**Abstract:** Im Rahmen der FahrzeuglackiererInnen-Ausbildung kann Virtual Reality (VR) zur Förderung des handlungsorientierten Erlernens von Techniken zur Durchführung von Kfz-Lackierarbeiten eingesetzt werden. Im Projekt HandLeVR wird dafür zunächst untersucht, welchen Beitrag VR-Technologien generell bieten, um Lernen in der Berufsausbildung zu ermöglichen bzw. wie diese Technologien didaktisiert werden können, um entsprechende Lernerfolge sicherzustellen. Auf Grundlage dieser Fragestellung wird eine VR-Lackierwerkstatt als virtuelle 3D-Lernumgebung mit Autoren-, Trainings- und Reflexionskomponente gestaltet, implementiert und im berufsbildenden Einsatz evaluiert.

**Keywords:** Virtual Reality, Handlungsorientiertes Lernen, Berufliche Bildung, Lackierausbildung

## 1 Projekt HandLeVR

Im Sinne der Kompetenzorientierung in der Berufsbildung ist es notwendig, häufig handlungsorientierte Lerneinheiten mit Lernfortschrittskontrollen durchzuführen. Gerade in Bereichen wie der Ausbildung von FahrzeuglackierInnen ist dies aus gesundheitlichen und ökologischen Gründen nicht in ausreichendem Maße sinnvoll. Durch den Einsatz Virtueller Realität (VR) könnte dieser Herausforderung effektiv begegnet werden. Im BMBF-geförderten Projekt HandLeVR[2] wird die Forschungsfrage untersucht, welchen konkreten Beitrag VR-Technologien bieten, um handlungsorientiertes Lernen in der Berufsausbildung zu ermöglichen und wie diese Technologien didaktisiert werden können, um entsprechende Lernerfolge sicherzustellen.

## 2 VR-Lackierwerkstatt

Zentrales Untersuchungsinstrument ist ein Lernsystem mit VR-Anteil – die VR-Lackierwerkstatt. Diese wird derzeit in einem agilen Entwicklungsprozess konzipiert, implementiert und evaluiert. Die VR-Lackierwerkstatt besteht, wie in Abbildung 1 visualisiert, aus drei Komponenten. Das Autorenwerkzeug wird zur Vorbereitung einer konkreten Lernhandlung vom Ausbilder bzw. der ausbildenden Institution genutzt, um

[1] Universität Potsdam, Institut für Informatik und Computational Science, August-Bebel-Str. 89, 14482, Potsdam, vorname.nachname@uni-potsdam.de
[2] HandLeVR wird gefördert vom Bundesministerium für Bildung und Forschung (BMBF), Förderkennzeichen: 01PV18002A, Laufzeit: 01.01.2019 - 31.12.2021

doi:10.18420/delfi2019_314

das allgemeine Szenario (Auftrag von Farbschichten) zu konkretisieren (z.B. durch Einbindung fertiger Templates für konkrete Lerneinheiten und/oder neuer 3D-Modelle).

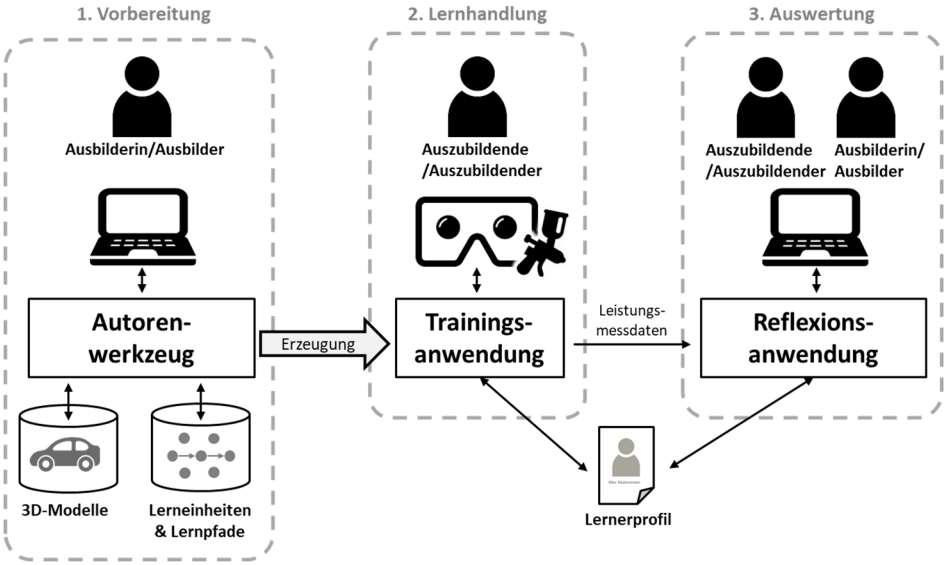



Abb. 1: VR-Lackierwerkstatt aus drei Komponenten inkl. einer VR-Trainingsanwendung

Die VR-Trainingsanwendung dient im berufsfachlichen Kern dem Training und der Qualitätskontrolle von Farbaufträgen auf 3D-Werkstücke. Auf der technischen Seite bestehen hier insbesondere Herausforderungen bei der Nachstellung der Physik des Farbauftrags (z.B. Farbpartikeldichte) und der Nachbildung der Lackierpistole sowohl in VR als auch als begreifbares, sensor-bestücktes Eingabegerät. Abbildung 2 zeigt einen aktuellen Einblick in die VR-Trainingsanwendung.

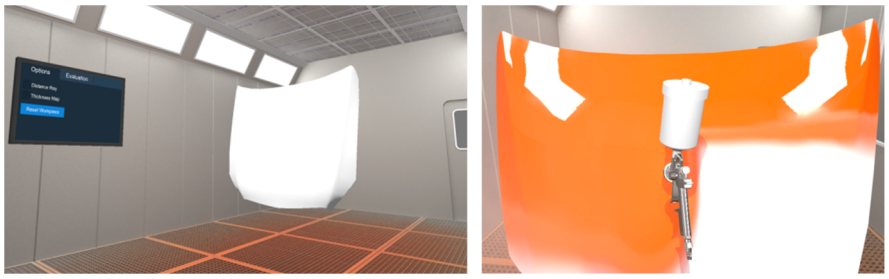


Abb. 2: In der Trainingsanwendung können Fahrzeugteile authentisch in VR lackiert werden.

In der abschließenden Reflexionsanwendung können die Lernleistungen mit Mit-Auszubildenden oder dem Ausbilder ausgewertet werden.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 307

# ILSA – ein integriertes Lernstil-Analysesystem für Moodle

Sven Judel[1], René Röpke[1], Nađa Žarić[1] und Ulrik Schroeder[1]

**Abstract:** Um Lernende in ihrer Lernumgebung besser zu erfassen und Lehrenden die Möglichkeit zu geben, ihre Lehre zielgruppengerechter zu gestalten, kann ein integriertes Lernstil-Analysesystem zur Verknüpfung von Lernstildaten mit Aktivitäts- und Leistungsdaten verwendet werden. In diesem Beitrag wird ein eben solches System für das Lernmanagementsystem Moodle vorgestellt und die Vorteile der unmittelbaren Integration in Moodle präsentiert.

**Keywords:** Lernstile, Learning Analytics, Lernumgebungen, Moodle

## Konzept und Implementierung

Zur besseren Erfassung von Lernenden innerhalb einer Lernumgebung können Lernstile (LS) nach dem *Felder-Silverman Learning Style Model* (FSLSM) [FB16] erhoben und mittels Learning Analytics mit den Aktivitäts- und Leistungsdaten der Lernenden korreliert werden. Ein erstes Konzept wurde in [ZRS18] vorgeschlagen und im Rahmen dieses Beitrags wird eine mögliche Umsetzung für das sehr weitverbreitete Lernmanagementsystem (LMS) Moodle[2] vorgestellt. Das integrierte Lernstil-Analysesystem (auf Englisch: *Integrated Learning Styles Analytics System*, kurz ILSA) basiert auf zwei Komponenten zur Identifikation und Korrelation von LS-Daten sowie Aktivitäts- und Leistungsdaten. Die erste Komponente bietet eine Implementierung des *Index of Learning Styles* (ILS) [SF19], die Lernende zur Identifizierung ihrer LS bearbeiten können. Die LS werden unmittelbar berechnet und in unterschiedlichen Visualisierungen für die Lernenden aufbereitet. Basierend auf ihren LS erhalten die Lernenden Informationen über die Charakteristika ihrer LS und Vorschläge, deren Umsetzung die eigenen Lernerfahrungen verbessern sollen.

Die zweite Komponente verknüpft die erhobenen LS-Daten mit Aktivitäts- und Leistungsdaten des LMS Moodle. Lehrende haben die Möglichkeit verschiedene LS-Gruppen bezüglich ihrer Aktivitäten zu untersuchen und zu vergleichen (siehe Abb. 2) Dazu werden in den in Abbildung 1 gezeigten Optionen die Aktivitäten des LMS ausgewählt, sowie die Visualisierungsform (z.B. Balken- oder Tortendiagramm) spezifiziert. Zudem kann der Zeitraum festgelegt werden, über welchen die Daten korreliert werden sollen.

[1] RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen, {judel, roepke, zaric, schroeder}@informatik.rwth-aachen.de
[2] https://moodle.org/, zuletzt abgerufen am 19.03.2019

doi:10.18420/delfi2019_301

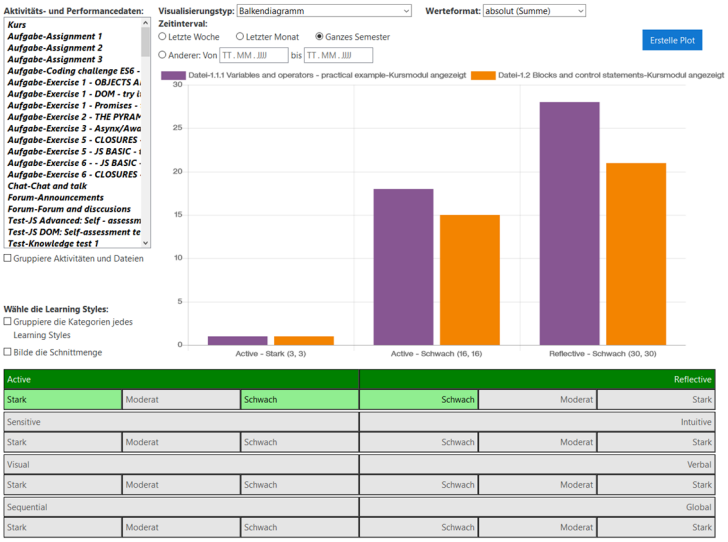



Abb. 1: Ansicht der Lehrenden auf die zweite Komponente

Lernende können die Analysekomponente verwenden, um sich mit den korrelierten Daten der Lernenden ihrer Lernstile zu vergleichen. Daraus können z.B. Schlüsse gezogen werden, welchen Materialien oder Lehreinheiten andere Lernende des eigenen Lernstils mehr Aufmerksamkeit gewidmet haben. Dies ergänzt die abstrakten Empfehlungen der ersten Komponente, um konkrete Indikatoren zur Verwendung bestimmter Lernmaterialien oder -aktivitäten.

Die Integration in das LMS Moodle hat essentielle Vorteile. Durch eine direkte Integration in das LMS unterschiedliche Datenquellen ausschöpfen (z.B. Notenbuch, Logdateien). Bestehende APIs erlauben den Zugriff auf LMS-Daten, sodass keine Einspeisung zusätzlicher Daten nötig ist. Ein Export von Rohdaten und Visualisierungen erlaubt die Weiternutzung der Daten in anderen Systemen. Weiterhin bleiben Lehrende und Lernende in einem gewohnten Umfeld, dem LMS, und müssen nicht in ein neues System wechseln, um die Dienste von ILSA zu verwenden. Die datenschutzrechtliche Grundlage zur Verwendung der Daten wird von Moodle zentral definiert und eingeholt.

## Literaturverzeichnis

[FB16] Felder, R. M.; Brent, R.: Teaching and Learning STEM: A Practical Guide, John Wiley & Sons, 2016.

[SF19] Soloman, B. A.; Felder, R. M.: Index of Learning Styles Questionnaire. https://www.webtools.ncsu.edu/learningstyles, abgerufen am 04.06.2019.

[ZRS18] Zaric, N.; Röpke, R.; Schroeder, U.: Concept for linking learning analytics and learning styles in e-learning environment. In: 10th International Conference on Educational and New Learning Technologies, Palma, S. 4822-4829, 2018.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 309

# Informatik Grundbildung im Alter – Ein eLearning Konzept

Svenja Noichl[1] und Ulrik Schroeder[1]

**Abstract:** „Ich möchte in Kontakt bleiben mit…" ist eine häufige Aussage, die in Tablet-Workshops für Seniorinnen und Senioren als Motivation zur Teilnahme geäußert wird. Das hier vorgestellte Konzept zur Vermittlung informatischer Grundkonzepte für Seniorinnen und Senioren, setzt thematisch an diesem Punkt an. Basierend auf Interesse, aktueller Bedeutung, Verfügbarkeit bereits existierender Materialien und Anknüpfungsmöglichkeiten an die Lebenswelt der Seniorinnen und Senioren wurden zu vermittelnde Kompetenzen ausgewählt.

**Keywords:** Informatik Bildung, Seniorinnen und Senioren, Kommunikation mit Informatiksystemen, Funktionsweise des Internets, Datenschutz und Datensicherheit

## 1 Motivation und Auswahl der Inhalte

Mit zunehmender Digitalisierung in allen Lebensbereichen werden grundlegende Informatikkompetenzen auch für ältere Erwachsene immer wichtiger. In Grundlagenworkshops zur Nutzung mobiler Geräte [vgl. NRS18] äußern die Teilnehmerinnen und Teilnehmer den Wunsch mit (entfernt lebenden) Familienmitgliedern oder Freunden in Kontakt zu bleiben als Motivation teilzunehmen. Neben der reinen Nutzungskompetenz entsprechender mobiler Geräte, kann die Kenntnis der Funktionsweise helfen, selbstständig mit der „neuen" Technologie umzugehen.

In einem zwei stufigen Verfahren wurden die Inhalte für vier Module unter dem Motto „Ich möchte Kommunizieren!" ausgewählt. Zuerst wurden nationale [vgl. z. B. GI08] und internationale [vgl. z. B. Cs11] Schulcurricula und Standards zusammengetragen. Basierend auf den Interessen der Zielgruppe (erhoben in einer Fragebogenstudie mit 123 Personen ab 50 Jahren) sowie Themen aktueller Relevanz in Deutschland, konnte ein Set an Informatik-Kompetenzen, für technikferne Erwachsene und Seniorinnen und Senioren, ermittelt werden. Anschließend wurden daraus, basierend auf Interesse, aktueller Bedeutung, Verfügbarkeit bereits existierender Materialien und Anknüpfungsmöglichkeiten an die Lebenswelt der Zielgruppe, eine Auswahl von Kompetenzen aus den Bereichen *Internet und Netzwerke*, *mobile Geräte und das Internet*, *Datenschutz und Datensicherheit*, *Kommunikation mit Informatiksystemen* sowie *offene Materialien und Urheberrecht,* getroffen. Die zu vermittelnden Inhalte aus diesen Bereichen werden in vier aufeinander aufbauenden Modulen strukturiert.

[1] RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, {noichl, schroeder}@informatik.rwth-aachen.de

doi:10.18420/delfi2019_245

## 2 Umsetzung der Module in einem eLearning Konzept




Abb. 1: Vier Module mit ihren Leitfragen und Zielen

Die in Abbildung 1 dargestellten Module werden in Form einer mobilen Android Applikation (kurz: App) umgesetzt. Angereichert werden die Inhalte der Module dabei mit Quizzen oder kleinen Aufgaben zur Anwendung des Erlernten. Auf diese Weise kann die App Lernwerkzeug und Messwerkzeug zur gleichen Zeit sein. Als Lernwerkzeug bietet sie die Möglichkeit die Lerninhalte auf multimediale, interaktive Weise zur Verfügung zu stellen. Als Messwerkzeug bietet sie die Möglichkeit zur Anbindung von Learning Analytics-Komponenten. In der Kombination kann so ein personalisiertes und an die individuellen Nutzerinnen und Nutzer angepasstes Lernerlebnis geschaffen werden. Dieses eLearning Konzept vereint drei wesentliche Komponenten in einer App, Inhaltskomponente (Darstellung der Inhalte), Testkomponente (Anwendung der Inhalte) und Learning Analytics-Komponente (Rückmeldung an Nutzende und Evaluation des Systems). Nach der technischen Umsetzung der hier vorgestellten Module müssen diese sowohl zur Verbesserung der Usability als auch zur Prüfung ihrer Effektivität in Bezug auf einen Zuwachs digitaler, informatischer Kompetenzen evaluiert werden.

Literaturverzeichnis

[Cs11] CSTA, K.: Computer science standards. Computer Science Teachers Association.

[GI08] GI, Gesellschaft für Informatik: Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I. LOG IN 28 (150/151), 2008.

[NRS18] Noichl, S.; Röpke, R.; Schroeder, U.: Adaptive Buttons für zielgruppengerechtes App Design. Mensch und Computer 2018-Workshopband, 2018.

# Inklusives Lernen mit barrierefreien Videos

**Ein Fallbeispiel**

Yasmin Patzer[1] und Niels Pinkwart[1]

**Abstract:** Lernen wird zunehmend durch den Einsatz digitaler Medien unterstützt, wobei die Heterogenität der Lerner*innengruppen zunimmt. Existierende Lernplattformen decken diese Heterogenität bisher nicht zufriedenstellend ab. Die inklusive Lernplattform LAYA (Learn As You Are) setzt hier an. Im Rahmen dieses Beitrags wird die Nutzbarkeit verschiedener Videoplayer in inklusiven Settings diskutiert. Inwieweit erfüllen existierende Player beispielsweise gängige Accessibility Guidelines? Im Anschluss zeigen wir anhand eines Fallbeispiels wie die Integration eines geeigneten Players in eine inklusive Lernplattform aussehen kann.

**Keywords:** Inklusion, inklusives eLearning, Barrierefreie Videoplayer, Barrierfreie Lernvideos

## 1 Motivation

Die Bedeutung von Videos für die Wissensvermittlung hat in den letzten Jahren deutlich zugenommen. Plattformen wie YouTube zeigen dies sehr eindrücklich. Auch in formaleren Lernszenarien finden sich zunehmend Videos als Medium zur Wissensvermittlung wieder. Mit Blick auf die Heterogenität von Lerngruppen bedeutet das, dass sowohl Inhalte, als auch Videoplayer inklusiv gestaltet werden müssen. González et al. [GMM15] formulieren Kriterien, die barrierefreie Videoplayer erfüllen sollten. Untersuchungen verschiedener Player kommen zu dem Ergebnis, dass die Mehrheit der Videoplayer nicht als barrierefrei bezeichnet werden kann [Wi18].

## 2 Die Videokomponente in LAYA anhand eines Fallbeispiels

Zu Beginn wurde in LAYA der HTML5 Videoplayer Plyr genutzt. Sein modernes schlankes Design sowie die Umsetzung von Accessibility Features wie Captions und guter Nutzbarkeit für Screenreader waren ausschlaggebend. Es zeigte sich schnell, dass der Funktionsumfang im Hinblick auf Barrierefreiheit nicht ausreichend ist. Dies deckt sich mit den Ergebnissen von Wild zur Barrierefreiheit verschiedener Videoplayer [Wi18]. Aktuell gibt es lediglich zwei barrierefreie Videoplayer, die als barrierefrei bezeichnet werden können, OzPlayer und AblePlayer [Wi18]. Der OzPlayer ist ein kommerzielles Produkt des Unternehmens AcessibilityOz und schied deshalb für die

[1] Humboldt-Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, patzer@informatik.hu-berlin.de, pinkwart@hu-berlin.de

doi:10.18420/delfi2019_325

Nutzung in LAYA aus. Der AblePlayer als OpenSource Produkt passt zur Philosophie von LAYA und wurde deshalb für die weitere Nutzung in der Lernplattform ausgewählt.

Aktuell befinden sich Materialien des Disabled and Self Employed (DSE) Projekts[2] in LAYA. Bei dem Projekt handelt es sich um ein EU-Projekt zur Selbstständigkeit von Menschen mit Behinderungen. Das Projekt hat videobasierte Lerneinheiten entwickelt, die zentrale Aspekte von Selbstständigkeit vermitteln. Der DSE-Kurs bietet die Möglichkeit für eine erste umfangreiche Nutzung des Videoplayers in LAYA. Im DSE-Kurs stehen die Lernvideos in fünf verschiedenen Sprachen (Deutsch, Spanisch, Italienisch, Niederländisch und Polnisch) zur Verfügung, wobei nicht alle Videos in allen Sprachen verfügbar sind. Die Nutzer*innen können in allen Sprachen zwischen „normalem“ und leichtem Sprachniveau wählen. Alle Videos umfassen Untertitel, die je nach Wunsch der Nutzer*innen zu- und abgeschaltet werden können. Darüber hinaus sind im Videofenster selbst Schaltflächen zum Beschleunigen oder Verlangsamen der Abspielgeschwindigkeit vorhanden. Weiterhin findet sich dort ein Button zum Zu- bzw. Abschalten des Transkripts. Für das Transkript findet ein Highligthing der jeweils aktuellen Position im Video statt, weiterhin kann automatisches Scrollen aktiviert werden. Das Fenster des Transkripts lässt sich beliebig platzieren, ebenso wie das zu- und abschaltbare Fenster für das parallel abgespielte Gebärdensprachvideo.

## 3 Ausblick

Im Rahmen eines Usability-Tests werden im Sommer 2019 anhand des videobasierten DSE-Kurses Rückmeldungen von heterogenen Nutzer*innen erhoben. Der Fokus liegt auf der Evaluation der Videokomponente, inklusive ihrer verschiedenen Unterstützungsmöglichkeiten. Aufgrund der vielen Unterstützungsmöglichkeiten besteht die Gefahr, dass es für Nutzer*innen zu unübersichtlich wird. Eine Adaptierbarkeit des Layouts nach den Präferenzen der Nutzer*innen könnte langfristig eine geeignete Lösung sein und entspräche dem restlichen Konzept der Lernplattform.

Literaturverzeichnis

[GMM15] González, M., Moreno, L., & Martínez, P. Approach design of an accessible media player. Universal Access in the Information Society, 14(1), 45–55, 2015.

[Wi18] Wild, G. (2018). The inaccessibility of video players. In (Miesenberger K., Kouroupetroglou G.): ICCHP 2018. Springer, Vol. 10896 LNCS, 47–51, 2018.

[2] Projektseite DSE: https://disabledandselfemployed.wordpress.com/

# Interaktive kollaborative Elemente aus einer domänenspezifischen Perspektive des Wirtschaftsunterrichts

Manuel Froitzheim[1] und Michael Schuhen[2]

**Abstract:** In diesem Posterbeitrag wird die Integration von domänenspezifischen Methoden zum kollaborativen Lernen mit mobilen Endgeräten in einem elektronischen Schulbuch dargestellt. Am Beispiel des wirtschaftswissenschaftlichen Unterrichtsinhalts „Nutzen und Grenznuten“ wird die Umsetzung in einer Lerneinheit für den Wirtschaftsunterricht am Ender der Sekundarstufe vorgestellt.

**Keywords:** Elektronisches Schulbuch, Kollaborativität, Interaktivität, ECON EBook

## 1 Einleitung

Ziel des Wirtschaftsunterrichtes ist, dass Schülerinnen und Schüler ökonomisch Denken und Handeln lernen. Im Wirtschaftsunterricht dominiert das gedruckte Schulbuch und somit eine textuelle Vermittlung ökonomischer Inhalte, deren Lernertrag im Vergleich zu dem Einsatz domänenspezifischer Methoden gering ist [We19]. Unter einer domänenspezifischen Methode wird eine Unterrichtsmethode verstanden, die das Erlernen des Denkens und Handelns in einem Fach unterstützt. So sind typische domänenspezifische Unterrichtsmethoden des Wirtschaftsunterrichts ökonomische Experimente, Planspiele oder Simulation, wohingegen für den Chemieunterricht das Schülerexperiment als fachliche Unterrichtsmethode im Zentrum steht.

Diese Methoden stellen in der Regel für die Lehrerinnen und Lehrer einen erheblichen Vorbereitungs- und Durchführungsaufwand dar, da Rollenkarten vorbereitet oder Entscheidungsdaten bei ökonomischen Experimenten eingesammelt werden müssen. Da insbesondere Planspiele und ökonomische Experimente auf Rechenkernen beruhen, müssen diese ferner vom Lehrer administriert werden. Durch die Integration dieser domänenspezifischen Methoden in ein Learning Management System oder in ein elektronisches Schulbuch werden sowohl der Vorbereitungsaufwand als auch der Durchführungsaufwand minimiert [No16].

Die kollaborativen domänenspezifischen Methoden wurden in das von den Autoren entwickelte elektronische Schulbuch ECON EBook (vgl. www.econ-ebook.de) integriert [FS15].

[1] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57068 Siegen, froitzheim@zoebis.de
[2] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57068 Siegen, schuhen@zoebis.de

doi:10.18420/delfi2019_337

## 2 Praxisbeispiel: Simulation zum Grenznutzen

Zu den zehn wichtigsten ökonomischen Prinzipien zählt das Konzept vom „Grenznutzen“ [MT18]. Dieses Konzept kann im Unterricht sowohl mathematisch hergeleitet als auch von den Lernenden mit Hilfe eines Textes erarbeitet werden. Beide Vermittlungsmethoden führen jedoch dazu, dass träges Wissen beim Lernenden entsteht und [We19] das Konzept meist nicht verstanden wird. Dies zeigt sich u.a. daran, dass die Schülerinnen und Schülern weder in der Lage sind, das Konzept auf andere Inhaltsfelder zu übertragen noch Anwendungsbeispiele mit Hilfe des Erlernten zu lösen.

Im Rahmen des zum Lerninhalt angebotenen ökonomischen Experiments erhalten die Lernenden in im ersten Schritt eine gesprochene Problemstellung. Nachdem die Lernenden die Tondatei zum Wandertag gehört haben, können sie ihren individuellen Preis für ein Glas Wasser in das mobile Endgerät eintragen. Der Kioskbesitzer bietet weitere Gläser an und der Lernende gibt erneut seine Zahlungsbereitschaft ein. Die Tondokumente sind jeweils auf die von den Lernenden eingegebene Zahlungsbereitschaft angepasst und in der Regel geben die Lernenden nach einigen Gläsern eine negative Zahlungsbereitschaft ein. Die Basis bildet ein Rechenmodell, dass die Entscheidungen der Lernenden interpretiert und jeweils eine passende Tondatei für die folgende Entscheidung aus einer Baumstruktur freigibt. Die Ergebnisse können mit Hilfe des volkswirtschaftlichen Modells für den Grenznutzen interpretiert und analysiert werden.

Literaturverzeichnis

[FS15] Froitzheim, M.; Schuhen, M.: Das ECON EBook als interaktives und multimediales elektronisches Schulbuch für den Ökonomieunterricht. In Pongratz, H.; Keil, R. (Hrgs.): DeLFI 2015 - Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik. Bonn: Köllen Druck+Verlag GmbH. S. 253-264. 2015.

[MT18] Mankiw, G.; Taylor, M.P.: Grundzüge der Volkswirtschaftslehre. Stuttgart; Schäffer-Poeschel, 7. Aufl.

[No16] Nosko, C.: Entwicklungschancen und Sackgassen. Das digitale Schulbuch kommt zu sich selbt. In (Schuhen, M.; Froitzheim, M.; Schuhen, K. (Hrsg.): Das Elektronische Schulbuch 2016. Fachdidaktische Anforderungen und Ideen Treffen auf Lösungsvorschläge der Informatik. LIT Verlag, Münster, 2016.

[We19] Weyland, M.: Experimentelles Lernen im Ökonomieunterricht – wirksam, effizient, unverzichtbar?. In: Zeitschrift für ökonomische Bildung, Ausgabe 9, 1-26. 2019.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 315

# Mit der Lupe unterwegs – eine spielbasierte Lernanwendung zu Sicherheit im Internet

René Röpke[1], Kathrin Larisch[1], Sven Schöbel[2] und Ulrik Schroeder[1]

**Abstract:** Der sichere Umgang mit IT-Systemen und dem Internet ist für Endnutzerinnen und Endnutzer ohne informatische Grundbildung eine große Herausforderung. Während die Nutzung an sich für viele Altersgruppen keine Schwierigkeiten darstellt, ist IT-Sicherheit ein sehr komplexes Thema mit dem sich nur selten angemessen auseinandergesetzt wird. Gerade Kinder im Alter von 9-12 Jahren lernen den Umgang mit Tablet und Smartphone sowie Apps und Web-Diensten im Handumdrehen. Was es bedeutet, sich *sicher* zu verhalten, vertiefen nur wenige. Dieser Beitrag präsentiert eine spielbasierte Lernanwendung zur Sicherheit im Internet. Als Detektiv bewegen sich die Spielerinnen und Spieler durch eine emulierte Browseroberfläche und entdecken Themen, wie den Aufbau von URLs und Phishing.

**Keywords:** Spielbasiertes Lernen, Lernspiel, IT-Sicherheit, Phishing, URL

Der Umgang mit IT-Systemen und dem Internet ist in der durch digitale Medien geprägten Welt für manche ein Kinderspiel, für andere eine große Herausforderung. Während Kinder den Umgang schnell erlernen und es als intuitiv empfinden, verhalten sie sich nicht unbedingt *sicher*. Hinzu kommt, dass auch ohne Wissen über sicheren Umgang mit IT-Systemen und dem Internet, diese Medien dennoch verwendet werden und Endnutzerinnen und Endnutzer sich dadurch unbewusst Risiken aussetzen.

Um den Zielgruppen verschiedenen Alters und ohne informatische Grundbildung den sicheren Umgang zu vermitteln, müssen grundlegende Inhalte und Fähigkeiten der Informatik und IT-Sicherheit zielgruppengerecht und motivierend aufbereitet werden. Digitale, spielbasierte Ansätze zeigen in diesem Bereich erste Erfolge [HAV16] und es existieren auch erste Prototypen für Kinder im Alter von 9-12 Jahren.

Das Lernspiel „Internet Hero“ der Universität Wien versetzt Spielerinnen und Spieler in eine fiktionale Welt, die das Internet repräsentiert. Die Spielerinnen und Spieler erlernen die technischen und sozialen Grundlagen der Internetnutzung durch eine Reihe von Mini-Spielen mit übergreifender narrativer Handlung. Erste Studien zeigen, dass die Grundgestaltung für die Zielgruppe (9-12 Jahre) angemessen und motivierend ist [Ka14].

Die vom Federal Bureau of Investigation (FBI) betriebene Plattform „Safe Online Surfing“[3] bietet für Kinder unterschiedlicher Altersstufen (nach Schulklassen sortiert)

[1] RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen, {roepke, larisch, schroeder}@informatik.rwth-aachen.de
[2] RWTH Aachen, Templergraben 55, 52062 Aachen, sven.schoebel@rwth-aachen.de
[3] https://sos.fbi.gov/, zuletzt abgerufen am 16.01.2019

doi:10.18420/delfi2019_217

Mini-Spiele zum Lernen über IT-Sicherheit. Die Einbettung in eine narrative Handlung fehlt jedoch und die Inhalte wirken somit zufällig präsentiert bzw. kontextfrei.

Die spielbasierte Lernanwendung dieses Beitrags baut auf den Erkenntnissen zur Gestaltung von Lernspielen [Ge03] auf und versetzt die Spielerinnen und Spieler mithilfe einer emulierten Browseroberfläche und dem Spielpartner „Detektiv Dot“ in eine risikofreie Spielwelt. Im ersten Mini-Spiel soll eine Phishing Academy besucht werden, bevor im Anschluss eine Reihe von ungelösten Kriminalfällen rund um das Thema Phishing zu lösen sind (siehe Abbildung 1).

Die spielbasierte Lernanwendung verwendet eine Lupe zum Entdecken von versteckten Inhalten und zur interaktiven Auseinandersetzung mit den Lerninhalten. Durch die emulierte Browseroberfläche werden die Spielerinnen und Spieler aus der echten Welt abgeholt und in die Spielwelt eingeführt. Anders als im echten Browser, können hier Fehler gemacht und Konsequenzen simuliert werden.

Die Anwendung ist webbasiert und unterstützt Multitoucheingaben. Sie wurde mithilfe des MTLG Frameworks [ELS17] entwickelt. Die Nutzung an Tablets oder im Browser eines Desktopgeräts wird empfohlen. Die modulare Gestaltung erlaubt die Einbettung weiterer Mini-Spiele und Inhalte.

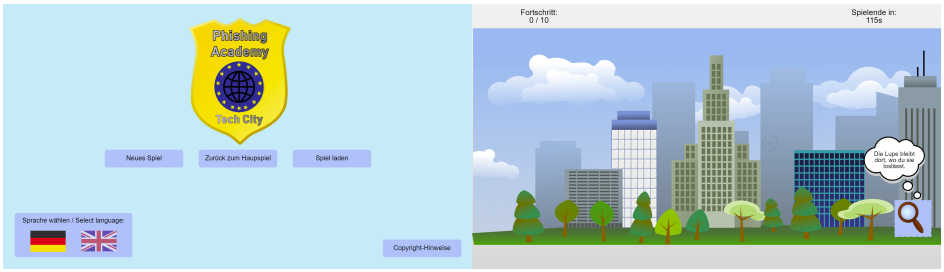



Abb. 1: Screenshots der spielbasierten Lernanwendung.

## Literaturverzeichnis

[ELS17] Ehlenz, M.; Leonhardt, T.; Schroeder, U.: Spielend leicht Lernspiele entwickeln - Ein Framework für Multitouch-Lernspiele. In (Igel, C. et al.): Bildungsräume 2017. Gesellschaft für Informatik, Bonn, S. 297-302, 2017.

[Ge03] Gee, J.P.: What video games have to teach us about learning and literacy. Computers in Entertainment (CIE), 1. Jg., Nr. 1, S.20, 2003.

[HAV16] Hendrix, M.; Al-Sherbaz, A.; Victoria, B.: Game based cyber security training: are serious games suitable for cyber security training? International Journal of Serious Games, 3. Jg., Nr. 1, S. 53–61, 2016.

[Ka14] Kayali, F. et al.: A case study of a learning game about the Internet. In: International Conference on Serious Games. S. 47–58, 2014.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 317

# Mit G/P/S durch die Welt der spielbasierten Lernanwendungen und Serious Games für IT-Sicherheit

René Röpke[1] und Ulrik Schroeder[1]

**Abstract:** Es gibt eine große Vielfalt an Serious Games und spielbasierten Lernanwendungen für unterschiedliche Themenfelder. Um einen Überblick über existierende Spiele für IT-Sicherheit zu erhalten, kann unter Verwendung einer Taxonomie eine Klassifizierung vorgenommen werden. Dieser Beitrag präsentiert das Ergebnis der Anwendung des G/P/S Modells auf eine in systematischer Literatur- und Produktrecherche ermittelten Menge von Serious Games, spielbasierten Lernanwendungen und Lernspielen. Es kann somit als Wegweiser für weitere Forschungsvorhaben dienen.

**Keywords:** Serious Games, Spielbasiertes Lernen, IT-Sicherheit, Taxonomie, Review

Die Interaktion mit IT-Systemen und dem Internet ist in der durch digitale Medien geprägten Welt für manche ein Kinderspiel, für andere eine große Herausforderung. Doch was intuitiv erscheint, ist es nicht zu gleich richtig – oder *sicher*. Spielbasierte Ansätze zur Vermittlung der notwendigen Kenntnisse und Fähigkeiten können *sichere Teilhabe* ermöglichen. Um einen Überblick über verfügbare spielbasierte Ansätze zur Vermittlung von Kenntnissen und Fähigkeiten über IT-Sicherheit zu erlangen, erfolgt auf Basis einer umfassenden Literatur- und Produktrecherche eine Klassifikation mithilfe des G/P/S Modells von Djaouti et al. [DAJ11].

Das G/P/S Modell versucht sowohl die Dimension „Game“ als auch die Dimension „Serious“ von spielbasierten Anwendungen zu analysieren und ist daher mehrdimensional. Es klassifiziert Spiele anhand drei wesentlicher Aspekte: „Gameplay“, „Purpose“ und „Scope“. Nach der Literatur- und Produktrecherche bildeten 39 verfügbare Anwendungen die klassifizierbare Ergebnismenge (siehe Tab. 1)

Der Aspekt „Gameplay“ oder Spielablauf versucht zu erfassen, wie das Spiel gespielt wird. Spiele, die Regeln zum Gewinnen des Spiels besitzen werden hier als „game-based“ klassifiziert. Gibt es keine expliziten Regeln zum Gewinnen, sondern bspw. einen Punktestand und Ranglisten, so sind diese Spiele „play-based“. Für spielbasierte Lernanwendungen und Serious Games über IT-Sicherheit gibt es sowohl Spiele mit dem Ziel zu gewinnen als auch Spiele, welche den Fokus mehr auf das Spielen des Spiels, als auf das reine Gewinnen legen.

In „Purpose“-Dimension des G/P/S Modells ist es für Spiele durchaus möglich mehrere Zwecke zu forcieren. So kann ein Spiel zum einen lediglich auf die Vermittlung von Inhalten abzielen, zum anderen kann es aber auch aktive Übungselemente besitzen,

[1] RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen, {roepke, schroeder}@informatik.rwth-aachen.de

doi:10.18420/delfi2019_295

sodass Verhalten erlernt und trainiert werden kann. 27 der 39 Beiträge wurden als „Message Broadcasting“ klassifiziert. In nur 15 Beiträgen wird explizites Training geboten, z. B. in der Bewertung von Phishing-URLs. Lediglich vier der ermittelten Spiele basieren auf dem Austausch von Wissen im kollaborativen Spiel. Diese Spiele sind jedoch keine digitalen Spiele, sondern eben Karten- bzw. Brettspiele.

| **Gameplay** | **# Ergebnisse** | **Scope** | **# Ergebnisse** |
|---|---|---|---|
| Game-based | 17 | Markt | |
| Play-based | 22 | Bildung | 30 |
| **Purpose** | | Berufswelt | 8 |
| Vermittlung | 27 | Entertainment | 1 |
| Training | 15 | Zielgruppe | |
| Austausch | 4 | Öffentlichkeit | 20 |
| | | Berufstätige | 8 |
| | | Studierende, Schüler | 18 |

Tab. 1: Ergebnisse der mittels des G/P/S Modells klassifizierten Beiträge

Hinsichtlich des „Scope“ wird zwischen Markt und Zielgruppe unterschieden. Erwartungsgemäß wurden die meisten Beiträge für den Bildungsbereich entwickelt bzw. dienen zur Weiterbildung, formal oder non-formal. 8 der 39 Beiträge sind dabei für Berufstätige. Die Zielgruppenverteilung zeigt, dass mehr als die Hälfte aller Beiträge für die Öffentlichkeit geeignet sind, aber auch 18 Beiträge explizit die Verwendung mit Schülerinnen, Schülern und/oder Studierenden befürwortet. Für manche Spiele werden explizite Hinweise zur Einbindung in die Lehre gegeben.

Schlussendlich zeigt die Klassifikation mithilfe des G/P/S Modells [DAJ11] die Vielfalt existierender Spiele im Bereich IT-Sicherheit. Während nur wenige Kollaboration und Austausch fördern, gibt es umso mehr Spiele zur Vermittlung von Inhalten. Der Fokus der Spiele liegt aufgrund des naheliegenden Bedarfs auf der Öffentlichkeit sowie Lernenden in Schule und Hochschule.

Im Rahmen des ERBSE („Enable Risk-aware Behavior to Secure End-users“) Projekts des Graduiertenkollegs NERD.NRW kann basierend auf dieser Arbeit die systematische Entwicklung eigener Anwendungen auf anderen existierenden Spielen aufgebaut werden. Dieser Beitrag wird durch ein Poster mit Details zur Klassifizierung und den Ergebnissen unterstützt. Das Poster wird auf der Projektwebseite[2] veröffentlicht.

Literaturverzeichnis

[DAJ11] Djaouti, D.; Alvarez, J.; Jessel, J.: Classifying serious games: the G/P/S model. In: Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches, ICI Global, S. 118–136, 2011.

[2] https://nerd.nrw/de/forschungstandems/erbse/, zuletzt abgerufen am 21.06.2019

# MTLG – Helping students and teachers creating and analysing simulations and games from a didactical and technical perspective

Thiemo Leonhardt[1], Matthias Ehlenz[2], René Röpke[2], Nadine Bergner[1] and Ulrik Schroeder[2]

**Abstract:** MTLG is a framework for the creation of animations, simulations and games that focuses on simple operation through multi-touch input, support for didactic implementation and the evaluation of learning data. The framework is completely web-based and therefore system and platform independent. Over several years, prospective computer science teachers and computer science students have used the framework in our e-learning labs and seminars. We have iteratively improved and enhanced the framework and its applications over several use cases in various computer science classes. The framework and all learning applications are available as open educational resources.

**Keywords:** framework, simulation, animation, serious games, teacher training, computer science education

## 1 Architecture

It is rather difficult to create animations, simulations and interactive gamified applications for specific learning goals, because an implementation of an interactive learning application requires in-depth specialist knowledge in many areas.

The general idea of the MTLG[3] framework is to offer a convenient toolkit for ambitious developers to produce their own learning games and simulations with reduced effort and without deep understanding of the inner workings of higher programming languages and server infrastructure. The resulting applications are platform independent and optimised for touch interaction.

Therefore, the framework consists of two major components. The first one is the gameframe and it can be interpreted as a scaffolding tool, managing dependencies, offering a build tool and live server as well as a pre-configured continuous integration pipeline for easy deployment to a web server. The second component is the actual core. It manages control flow and provision of further modules as the framework can be extended by including modules for specific tasks and purposes. The framework as well as all modules are published under an open license as open educational resources. The

[1] Professur für Didaktik der Informatik, Technische Universität Dresden, Germany
[2] The Learning Technologies Research Group, RWTH Aachen University, Germany
[3] https://gitlab.com/mtlg-framework, last accessed on March 21st, 2019

doi:10.18420/delfi2019_331

target group is mainly technically interested teachers as well as prospective teachers in their studies. The core functionality and the most important modules are as follows:

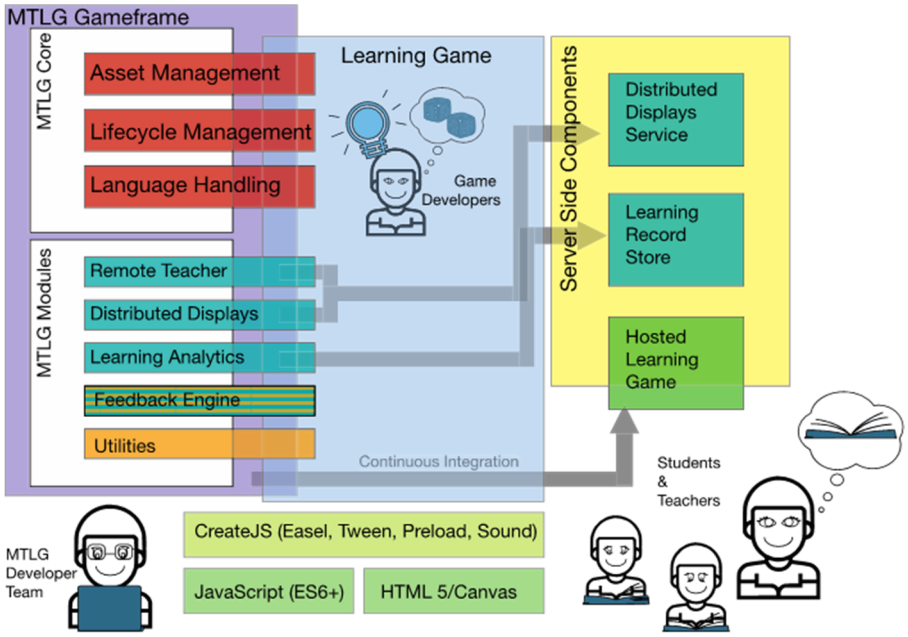



Abb. 1: Architecture

In the future, the framework will provide further support for didactically meaningful planning and implementation of a collaborative learning application.

## Bibliography

[Eh18] Ehlenz, M. et al.: The Lone Wolf Dies, the Pack Survives?: Analyzing a Computer Science Learning Application on a Multitouch-Tabletop. In (Joy, M.; Ihantola, P. Eds.): Proceedings, 18th Koli Calling Conference on Computing Education Research. Koli Calling 2018 November 22-25, 2018, Koli, Finland. The Association for Computing Machinery, New York, 2018; pp. 41–48.

[Le18] Leonhardt, T. et al.: Kollaborative Multiuser-Serious-Games für elektronische Schulbücher Lernspiele in der {Informatik}. In (Schuhen, M.; Froitzheim Eds.): Das elektronische Schulbuch 2017. Fachdidaktische Anforderungen und Ideen treffen auf Lösungsvorschläge der Informatik. LIT, Berlin, 2018; pp. 145–156.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 321

# Nachhaltiges und Aktivierendes Lernen durch eigene Erklärungen (NALE) mit geteilten, digitalen Portfolios

Uwe Lorenz[1] und Tatjana Reinstorf[2]

**Abstract:** Durch aktuelle Fortschritte in der Softwareergonomie muss das Gestalten von Medien- und Softwareprodukten nicht mehr professionellen Spezialisten überlassen werden. Wir regen eine digitale „Erklärproduktorientierung“ in Unterricht und Lehre an, in Verbindung mit digitalen Portfolios, die geteilten Content beinhalten. Unsere Methode „Nachhaltiges und Aktivierendes Lernen durch eigene Erklärungen“, genannt NALE, die an der Stephen-Hawking-Schule, einem SBBZ[3] mit dem Förderschwerpunkt körperliche und motorische Entwicklung in Neckargemünd auch erprobt wurde, erweitert aktivierende und selbstorganisierte Lehr- und Lernmethoden durch die Produktion von digitalen Erklärmaterialien. „Nur was ich anderen erklären kann, habe ich selbst verstanden.“ Es hebt den Lernenden in eine aktive Position und macht ihn zu einem kreativen Gestalter. Dies ist in einem Umfeld von allgegenwärtigen digitalen Endgeräten und künstlicher Intelligenz von zentraler Bedeutung zum einen für nachhaltiges und flexibles Begreifen und zum anderen für humanistische Bildungsziele.

**Keywords:** konstruktivistischer, handlungsorientierter und produktorientierter Unterricht, digitale Erklärmaterialien, Projektunterricht mit digitalen Hilfsmitteln, Lernen durch Lehren (LdL)

Durch die neuen Gestaltungsmöglichkeiten mit aktuellen digitalen Werkzeugen können handlungsorientierte, konstruktivistische Unterrichtsmethoden wie „Lernen durch Lehren“ [Ma02] „erklärproduktorientiert“ weiterentwickelt werden. Das kreative Produzieren von Erklärungen erfordert die Fähigkeit, die Bedeutungen im Stoff zu erkennen und sich in die Vorstellungswelt des Publikums oder des möglichen Nutzerkreises hinein zu versetzen. Anderen Personen klar strukturiert aber auch kreativ und in knapper Form einen Sachverhalt verständlich zu präsentieren, mithin Grundkonzepte nachhaltig zu begreifen und somit auch ähnlich gelagerte Probleme in diversen Situationen wiedererkennen und lösen zu können, sind Schlüsselkompetenzen im 21. Jahrhundert („21st century skills“ [Re17]). Dies wird durch gegenwärtige Entwicklungen in der künstlichen Intelligenz und der Digitalisierung im Allgemeinen noch verstärkt.

Nach [RKB19] hält nahezu die Hälfte der Schülerinnen und Schüler YouTube-Videos für schulische Belange für wichtig bis sehr wichtig. Weiterhin gaben viele Befragte an, dass sie mit YouTube motiviert werden, selbst künstlerisch aktiv zu werden.

[1] SRH Stephen-Hawking-Schule, Im Spitzerfeld 25,69151 Neckargemünd, uwe.lorenz@srh.de
[2] SRH Stephen-Hawking-Schule, Im Spitzerfeld 25,69151 Neckargemünd, tatjana.reinstorf@srh.de
[3] Sonderpädagogisches Bildungs- und Beratungszentrum

doi:10.18420/delfi2019_345

Als zentrales Element des Lernprozesses möchten wir das kreative Erschaffen von eigenen Erklärprodukten setzen. Dabei sind natürlich Voraussetzungen zu beachten. Der Unterricht selbst ist in vier Phasen aufzuteilen und erweitert die bekannte „Think-Pair-Share-Methode“ [LF17] hin zu einem „Think-Pair-***Do***-Share“:

1. In der ersten Phase eignen sich die SuS notwendiges Wissen an, z.B. mittels Frontalunterricht oder selbstorganisiertem Lernen, „SOL“.
2. In der zweiten Phase tauschen sich die Lernenden mit Ihrem Partner über die zentralen Begriffe aus oder holen sich von den Lehrkräften Feedback ein.
3. In der dritten Phase gestalten und produzieren SuS eigene Erklärmaterialien (z.B. PowerPoint-Präsentationen, Erklärvideos, interaktive Arbeitsblätter, Tests…) mit Blick auf das künftige Teilen des Inhalts.
4. In der vierten Phase stellen die Lernenden ihre Erklärungen den anderen vor, sichern die Produkte in ihrem digitalen Portfolio oder veröffentlichen diese auf digitalen Plattformen.

Es wechseln sich Arbeits- und Feedbackphasen ab. In den Phasen 1 und 3 werden im eigenen Tempo Inhalte erarbeitet, in den Phasen 2 und 4 wird im Wesentlichen evaluiert und reflektiert. Die sozialen Mechanismen der Bewertung und Selektion von Inhalten steuern die Entwicklung des Contents und der Lehre anforderungsgemäß. Plattformen, die solchen erklärproduktorientierten Unterricht stützen wollen, müssen den Teilnehmerinnen und Teilnehmern zum Einen ermöglichen, sich selbstorganisiert Wissen anzueignen und zum Anderen erlauben, digitale Portfolios bzw. Sammlungen mit den eigenen oder fremden Erklärprodukten anzulegen.

NALE ermöglicht Lernenden mit erhöhten Fehlzeiten eine produktive Beschäftigung mit schulischen Inhalten, langsamere oder gehandicapte SuS in inklusiven Settings können mit Hilfe der geschaffenen Materialien auch Inhalte nachholen. Motorische Handicaps können durch rechnergestützte Hilfsmittel kompensiert werden.

## Literaturverzeichnis

[LF17] Lehrerfortbildung BW, Individuelle Förderung, Unterrichtsgestaltung, Methodenblätter, https://lehrerfortbildung-bw.de/st_if/bs/if/unterrichtsgestaltung/methodenblaetter/thinkpairshare.html, 2017, Stand: 23.3.2019

[Ma02] Jean-Pol Martin (2002): Lernen durch Lehren (LdL). Die Schulleitung - Zeitschrift für pädagogische Führung und Fortbildung in Bayern. Heft 4: 3-9.

[Re17] Red, Toby : Entwickeln von 21st century skills mit dem I-Pad, https://integrate2learn.de/2017/02/11/21st/, 2017, Stand 23.3.2019

[RKB19] Rat für Kulturelle Bildung, Studie „Jugend/YouTube/Kulturelle Bildung. Horizont 2019“.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 323

# Open Digital Lab for You (DigiLab4U)

## Entwicklung und Pilotierung eines mediendidaktischen Designs für laborbasiertes Lehren und Lernen

Anke Pfeiffer[1] und Dieter Uckelmann[2]

**Abstract:** Die Digitalisierung in Bildung und Forschung ermöglicht neue Formen der standortübergreifenden Vernetzung von Laborinfrastrukturen und kann gleichermaßen Studierenden und Forschenden praktische Handlungs-, Lern- und Forschungsfelder bieten. Das BMBF geförderte Projekt Open Digital Lab for You[3] (DigiLab4U) verfolgt in diesem Sinne das Ziel Studierenden, Lehrenden und Forschenden in naher Zukunft standortunabhängig praxisnahe, digitalisierte und vernetzte Laborumgebungen bereitzustellen. Das nachfolgend dargestellte laborbasierte didaktische Design wurde im Sommersemester 2019 in einer Lehrveranstaltung im Studiengang Bachelor Informationslogistik an der HFT Stuttgart implementiert. Hybride Laborlernumgebungen werden dabei mit der Methode des Team Labor Coachings (LabTC) didaktisch angereichert und mit Learning Analytics Elementen ergänzt.

**Keywords:** Laborbasiertes lernen und lehren, hybride Lernumgebung, digitales Laborteam Coaching, Learning Analytics

Um angehende Ingenieurinnen und Ingenieure erfolgreich auf den wissenschaftlichen und praktischen Umgang mit Themen aus dem Bereich „Internet of Things“ und „Industrie 4.0“ vorzubereiten, benötigen sie neben theoretischem Fachwissen, berufspraxisrelevante Fähigkeiten und Kompetenzen, die gerade in laborbasierten Experimenten gut erworben werden können [FR05], [Te16]. Neben dem Erwerb klassischer Fachkompetenz, lassen sich in Laborexperimenten auch fachübergreifende Kompetenzaspekte für unterschiedliche, berufliche Praxissituationen adressieren. Im hier vorgestellten didaktischen Design zeigt sich der Lernerfolg zum einen unmittelbar in den laborbasierten Übungen und zum anderen in der semesterbegleitenden Dokumentation der Ergebnisse im LMS Moodle. Da im Ingenieurstudium sowohl fachliche Kompetenzen als auch überfachliche Kompetenzen, wie Teamfähigkeit, Kommunikationsfähigkeit oder Zeitmanagement berücksichtigt werden sollen, orientiert sich das mediendidaktische Design für die Durchführung der laborbasierten Phasen im Studiengang IL an der Methode des Team Labor Coachings (LabTC) nach Jödicke und Sum [JS07]. Dabei werden die Phasen des LabTC Konzeptes (Erarbeitung, Vertiefung, Auswertung) um digitale

[1] Hochschule für Technik Stuttgart, Studiengang Informationslogistik, Schellingstraße 24, 70174 Stuttgart, anke.pfeiffer@hft-stuttgart.de
[2] Hochschule für Technik Stuttgart, Studiengang Informationslogistik, Schellingstraße 24, 70174 Stuttgart, dieter.uckelmann@hft-stuttgart.de
[3] Projektkonsortium: Hochschule für Technik Stuttgart (HFT), Bremer Institut für Produktion und Logistik (BIBA), Institut für Wissensmedien der Universität Koblenz-Landau (IWM), RWTH Aachen,Universität Parma.

doi:10.18420/delfi2019_310

Elemente erweitert, mit dem Ziel das Erfassen der fachlichen und überfachlichen Lern- und Lehrprozesse zu analysieren und bei Bedarf zu optimieren. Zu diesen digitalen Elementen zählen unterschiedliche Ressourcen, die über die Lernplattform Moodle bereitgestellt werden, wie digitale Skripte und Handreichungen, Videos, Tests, Online-Feedback, Forenbeiträge und Aufgabenstellungen. Darüber hinaus sollen zukünftig laborbasierte VR/AR Daten ebenfalls in die Lernanalyse mit einfließen. Die mit Hilfe der Lernplattform erhobenen Daten werden aufbereitet und Studierenden und Lehrenden im Semesterlauf regelmäßig als Feedback zum individuellen Lernstand, zu den durchgeführten Aktivitäten und zur Nutzung der bereitgestellten Lernressourcen auf der Lernplattform bereitgestellt [LE17]. Gerade für den Team- und Coachingprozess erweist sich die digitale Erweiterung als hilfreich, denn sie liefert Lehrenden und Studierenden Feedback zu den laborbasierten Lernaktivitäten und der Lernperformance. Da die selbstgesteuerten Lernphasen im Semesterverlauf zunehmen, ist es für Lehrende essenziell zu wissen, wie aktiv die Studierenden die bereitgestellten Ressourcen und Lernaktivitäten nutzen, um beispielsweise frühzeitig zu erkennen, ob grundlegende Inhalte bearbeitet wurden, ob mögliche Probleme in der Bearbeitung bestehen oder ob zusätzliche Maßnahmen erforderlich werden. Für Studierende werden ausgewählte LA-Daten genutzt, um ihnen ihre Studienleistung und ihre Mitwirkung innerhalb des Lernteams zu visualisieren. Die Ergebnisse der Analyse fließen im Coaching mit ein, um den laborbasierten Lernprozess zu besprechen und eine weitergehende Reflektion über gezielte Coachingfragen anzuregen. Die Learning Analytics Daten werden vor dem pädagogischen Hintergrund des LabTC-Ansatzes erhoben und interpretiert. Neben der Unterstützung der Lern- und Lehrprozesse dienen die LA-Daten auch zur Weiterentwicklung des mediendidaktischen Designs. Um die Wirksamkeit des Ansatzes zu überprüfen, werden ausgewählte Lehrveranstaltung, die LabTC einsetzen, zyklisch mit den folgenden Instrumenten evaluiert: Fragebogen, Teaching Analysis Poll Evaluation (TAP) und Learning Analytics Daten.

## Literaturverzeichnis

[FR05] Feisel, L.D.; Rosa, A. J.: The Role of the Laboratory in Undergraduate Engineering Education. Journal of Engineering Education 94, pp. 121–130, 2005.

[JS07] Jödicke, B.; Sum, J.: Team Labor Coaching (LabTC). Geschäftsstelle für Hochschuldidaktik an Fachhochschulen in Baden-Württemberg (GHD). 7. Tag der Lehre. Hochschule Biberach, 2007.

[LE17] Leitner, P.; Ebner, M.: Learning Analytics in Hochschulen. In (Erpenbeck, J.; Sauter, W., Hrsg.): Handbuch Kompetenzentwicklung im Netz: Bausteine einer neuen Welt. Schäffer-Poeschel Verlag, Stuttgart, pp. 371-383, 2017.

[Te16] Tekkaya, A. E.; Wilkesmann, U.; Terkowsky; Pleul, C.; Radtke, M.; Maevus, F. (Hrsg.): Das Labor in der ingenieurwissenschaftlichen Ausbildung. Zukunftsorientierte Ansätze aus dem Projekt IngLab: acatech Studie. Deutsche Akademie der Technikwissenschaften. Herbert Utz Verlag GmbH, München, 2016.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 325

# Verbesserung der Lehrpraxis durch die Systematisierung erprobter Lösungen: Die Plattform Patternpool

Mareike Bartels[1], Eileen Lübcke[1], Gabi Reinmann[1], Finja Huckfeldt[2] und Ivo van den Berk[2]

**Abstract:** Auf der Plattform Patternpool werden hochschulische Lehr-Lernarrangements in Form von Patterns einfach zugänglich gemacht. Lehrende können Patterns bedarfs- und problemorientiert suchen, rezipieren und im besten Fall in der eigenen Lehrpraxis anwenden (Open Content). Hochschulbildungsforschende können die Lehr-Lernarrangements in Bezug auf verschiedene Fragestellungen analysieren, vergleichen und daraus Erkenntnisse für die Hochschuldidaktik generieren (Open Research). Patternpool zielt somit auf den unmittelbaren Transfer von erprobten Lösungen in Lehrveranstaltungen sowie auf den mittelbaren Transfer durch hochschuldidaktische Beratungen, Workshops und Forschung ab.

**Keywords:** Educational Patterns, Didaktische Entwurfsmuster, Hochschuldidaktik, Lehr-Lernforschung, Lehrpraxis, Lehrqualität, Open Access, Open Content, Open Research

Die Verbesserung der Lehrqualität steht ganz oben auf der hochschulpolitischen Agenda. Lehrinnovationen werden gefordert und gefördert und schlagen sich in vielfältigen Lehr-Lernarrangements nieder, von denen man annimmt oder inzwischen weiß, dass sie geeignet sind, verschiedenen Herausforderungen wie Heterogenität, kritische Übergangsphasen, systematische Defizite etc. in verschiedenen Fächern erfolgreich zu begegnen. Viele innovative Lehrprojekte wissen allerdings untereinander oft wenig von den Ergebnissen und Erfahrungen der anderen, da aktuell keine gemeinsame Plattform zum Austausch von Erfahrungswissen gepflegt wird [SB16].

Mit OPTion (Open Pattern Tool for Higher Education Research and Practice) entsteht eine Plattform zur Dokumentation und nachhaltigen Nutzung von erprobten Lehr- und Lernarrangements. Mit OPTion wird ein Werkzeug im Sinne einer technischen Open-Access-Infrastruktur (weiter-)entwickelt und der Öffentlichkeit zugänglich gemacht. Dabei wird bewährtes Handlungswissen aus der Lehre nach dem Pattern-Ansatz [Al77] systematisiert und für andere Lehrende bereitgestellt.

Durch das Verfassen von Entwurfsmustern (Patterns) soll das Wissen über erprobte Lösungen für bestimmte Probleme zugänglich und nachvollziehbar gemacht werden. Der aus der Architektur stammende Pattern-Ansatz bietet sich zum Einsatz in der

[1] Universität Hamburg, Hamburger Zentrum für Universitäres Lehren und Lernen (HUL), Schlüterstr. 51, 20146 Hamburg, mareike.bartels@uni-hamburg.de, eileen.luebcke@uni-hamburg.de, gabi.reinmann@uni-hamburg.de
[2] Hochschule Emden/Leer, Campusdidaktik, Constantiaplatz 4, 26723 Emden, finja.huckfeldt@hs-emden-leer.de, ivo.vandenberk@hs-emden-leer.de

doi:10.18420/delfi2019_297

Hochschuldidaktik an, denn auch in der Lehre gibt es wiederkehrende Herausforderungen, für die erprobte Lösungen bestehen.

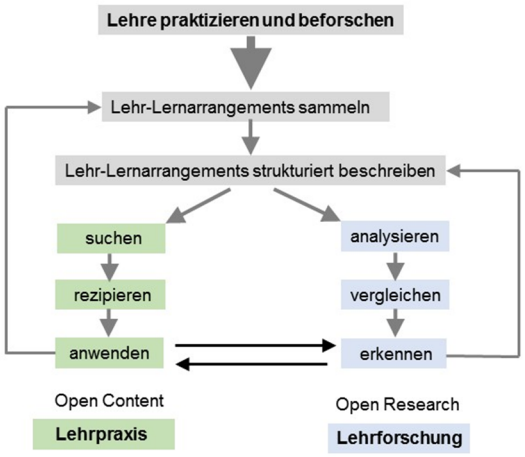



Abb. 1: Erstellung und Nutzung von Patterns auf Patternpool

Die systematische Erfassung und Darstellung von Entwurfsmustern erfolgt auf Patternpool entlang folgender Kategorien: Problem/Grund, Lösung/Maßnahme, Folgen/Wirkungen, Kräfte, Kontext und Details. Nutzer_innen können für ihren Lehrkontext relevante Patterns finden und für die eigene Praxis adaptieren. Als Autor_innen können Lehrende ihre eigene erprobte Lehrpraxis dokumentieren und anderen zur Verfügung stellen. Über die Plattform wird so ein Beitrag zur nachhaltigen Qualitätssicherung und -entwicklung in der Lehre geleistet.

## Literaturverzeichnis

[Al77] Alexander, C. et. al.: A Pattern Language. Towns, Buildings, Construction, Oxford University Press, New York, 1977.

[SB16] Schultes, K.; van den Berk, I.: Dokumentation und Distribution erprobter Lösungen – das Pattern-Pool-Tool „P2T“. In van den Berk, I. et.al. (Hrsg.): Studierfähigkeit. Theoretische Erkenntnisse, empirische Befunde und praktische Perspektiven, Universitätskolleg-Schriften, Bd. 15, S. 97-111, Universität Hamburg, Hamburg, 2016.

[Ko14] Kohls, C.: The Theories of Design Patterns and their Practical Implications exemplified for E-Learning Patterns. Dissertation. Katholische Universität Eichstätt-Ingolstadt, Eichstätt, 2014.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 327

# Quantitative Methodenkompetenzen im Praxissemester in Lehramtsstudiengängen

**Didaktisches Konzept eines ergänzenden Online-Angebots**

Till Massing[1], Natalie Reckmann[1], Benjamin Otto[2], Alexander Blasberg[1], Christoph Hanck[1] und Michael Goedicke[2]

**Abstract:** Das Praxissemester ist ein integraler Bestandteil in allen Lehramtsstudiengängen in Nordrhein-Westfalen. Dieses wird von der universitären Seite begleitet durch Seminarmodule, in denen eigene forschungsbezogene Abschlussberichte verfasst werden sollen. An vielen Universitäten ist jedoch eine Methodenausbildung nicht fest verankert. Um diese Lücke zu schließen wurde an der Universität Duisburg-Essen ein vollständig Online-basierter Selbstlernkurs entworfen. Das Konzept dieses Kurses und die Auswertung der angefallenen Daten ist der Fokus dieses Posters.

**Keywords:** Selbstlernkurs, Lehramtsstudiengänge, Praxissemester

## 1 Einleitung

Studierende, welche das Praxissemester in Lehramtsstudiengängen zu absolvieren haben, sollen Ihre Beobachtungen anhand eines Forschungsgegenstandes evaluieren. Dabei eignen sich sowohl qualitative als auch quantitative Herangehensweisen. Bei zweiter stehen statistische Auswertungen der Beobachtungen im Vordergrund.

Dies ist curricular durchaus eine gewisse Herausforderung, da Lehramtsstudierende verschiedener Fachrichtungen oft keine ausreichende statistische Vorbildung besitzen. Studiengänge, in denen Methodenkurse nicht verpflichtend sind, oder in denen sogar die Wahlmöglichkeit eines Methodenkurses fehlt, unterstützen die Studierenden nicht angemessen, wodurch diese und deren Betreuer vor, während und nach dem Praxissemester eine zusätzliche Beanspruchung erfahren.

Aus diesem Grund wurde an der Universität Duisburg-Essen im letzten Jahr ein Online-Selbstlernkurs mit einem Workload von ca. 30 Stunden für quantitative Methodenkompetenzen etabliert, welcher Studierende zur Vor- und Nachbereitung ihrer Forschungsfrage unterstützt. Dieser Online-Kurs wird auch von einigen Dozierenden des Begleitseminars zum Praxissemester via blended-learning Elementen genutzt.

[1] Universität Duisburg-Essen, Lehrstuhl für Ökonometrie, Universitätsstr 12, 45117 Essen, vorname.nachname@vwl.uni-due.de
[2] Universität Duisburg-Essen, paluno, Gerlingsstr 16, 45127 Essen, vorname.nachname@paluno.uni-due.de

doi:10.18420/delfi2019_300

## 2 Didaktisches Konzept

Der Online-Selbstlernkurs „Quantitative Methodenkompetenzen“ ist auf der Onlineplattform moodle der Universität Duisburg-Essen angesiedelt und für alle Studierenden von Lehramtsstudiengängen der Universität Duisburg-Essen frei zugänglich. Nach einer Aufbauphase des Kurses wird dieser seit dem Wintersemester 2018/2019 in einer ersten Testphase von Lehramtsstudierenden im Praxissemester erstmalig genutzt.

Der Kurs besteht aus mehreren integrativen Elementen, die die Studierenden eigenverantwortlich nutzen können, um sich quantitative Methodenkompetenzen anzueignen. Dabei wird der Kurs gegliedert in verschiedene statistische Themen, die teilweise aufeinander aufbauen und eine hohe Relevanz für die spezifischen Forschungsfragen für Lehramtsstudierende im Praxissemester aufweisen, wie zum Beispiel die Durchführung von Hypothesentests. Diese Kursstruktur gewährleistet einen Einstieg der Studierenden in den Online-Selbstlernkurs nach individuellen bereits erworbenen statistischen Kompetenzen.

Zu jedem einzelnen Thema werden im Online-Selbstlernkurs PDF-Präsentationsfolien bereitgestellt, die die Studierenden zum eigenverantwortlichen Lernen des jeweiligen Themas nutzen können. Eine Verzahnung mit Lehrvideos zu jedem Thema fördert dabei das audiovisuelle Lernen. Eine Prüfung des individuellen Lernfortschritts erfolgt über eine Vernetzung mit der Online-Lernplattform JACK, auf der die Studierenden zu jedem Thema Übungsaufgaben absolvieren können. Diese Übungsaufgaben nutzen den R-Commander, ein zusätzliches Tool des statistischen Programmpakets R, mit dem die Studierenden den Umgang mit quantitativen Daten erlernen um gleichzeitig ihre individuellen quantitativen Methodenkompetenzen zu festigen. Der R-Commander bietet den Vorteil, dass die Studierenden, im Gegensatz zu R, keine Programmierkenntnisse vorweisen oder erlernen müssen, sondern über eine Menüführung zum jeweiligen statistischen Konzept geleitet werden. Zum Erlernen des R-Commanders wurde dieser auf der Lernplattform JACK nachgebaut.

Zusätzlich enthält der Online-Selbstlernkurs „Quantitative Methodenkompetenzen“ ein Diskussionsforum, in dem die Studierenden eventuelle offene Fragen stellen können.

## 3 Evaluation

Zum jetzigen Zeitpunkt (März 2019) liegen noch keine finalen Evaluationsergebnisse vor, diese werden jedoch mit Ablauf dieses Semester erwartet. Die Kursnutzerzahl beträgt 66 Teilnehmer, sodass mit aussagekräftigen Beobachtungen gerechnet werden kann, welche auf dem Poster präsentiert werden.

# Promoting Competencies of the Future

## The Role of University-integrated Makerspaces

Bonny Brandenburger[1] [2] and Gergana Vladova[3] [4]

**Abstract:** An increasing interaction between man and machine in all areas of society leads not only to new job profiles but also initiates a discussion about the needed digital competencies in the future. The European Commission's Joint Research Centre has developed in a multi-stakeholder approach an initial framework to describe and measure the needed competencies such as problem solving ability, capacity of teamwork and programming skills [CVP17]. It is expected that the transfer of technical, occupation-oriented and personality-building competencies alongside subject-related knowledge will become increasingly important in higher education [SV16]. In order to avoid inequalities among students -those with and without so-called digital competencies - new learning approaches are needed. One possible approach towards the enhancement of digital competencies can be found within the maker movement.

**Keywords:** Higher education, digital competencies, maker movement, university-integrated makerspaces, collaborative and hands-on learning

## What role can making play in higher education?

Using digital desktop tools, makers create new products and prototypes. It is part of their philosophy to share ideas and drafts within the global online community. Makers use so-called makerspaces as working environment, which provide access to the latest technical infrastructure. In addition, the working environment of makerspaces enables an interdisciplinary and creative exchange of knowledge. Over the last twenty years, several public libraries, enterprises, schools and universities in the USA and Europe have set up makerspaces with the particular aim of promoting innovation [Fr16]. In numerous reports 'making' has been identified as an important trend both for universities and for the education system [Jo15] [Co17]. Makerspaces have been described as a key technology for the further development of universities [HD16]. For this reason, a number of projects such as "FAB101"[5] or "OD&M Alliance"[6] have been initiated to investigate the potential of digital fabrication in higher education. Nonetheless, there are very few empirical

[1] University of Potsdam, Department Business Informatics, Karl-Marx-Straße 67, 14482 Potsdam, bonny.brandenburger@uni-potsdam.de
[2] Weizenbaum-Institute, Hardenbergstraße 32, 10623 Berlin
[3] University of Potsdam, Department Business Informatics, Karl-Marx-Straße 67, 14482 Potsdam, gergana.vladova@wi.uni-potsdam.de
[4] Weizenbaum-Institute, Hardenbergstraße 32, 10623 Berlin
[5] FAB101 funded by the BMBF, Link: https://fab101.de/#project-summary-in-english
[6] OD&M Alliance co-funded by the Erasmus+,Programme of the EU, Link: https://odmplatform.eu/

doi:10.18420/delfi2019_315

analyses on concrete learning outcomes and competence development via making [ES17]. Taking into account the European Commission's definition of digital competencies [Fe12] and existing studies in the field of making and learning, making can be seen a promising learning approach for the enhancement of digital competencies [PGJ17]. However, appropriate methods and tools to develop competencies and to monitor learning outcomes in making are still missing. The contribution of this poster adds value at this point demonstrating a conceptual approach for the competence development in university-integrated makerspaces. Concretely, the link between the Digital Competence Framework for Citizens [CVP17], making and formal academic learning by the example of 3D printing workshops[7] will be established.

Bibliography

[CVP17] Carretero, S., Vuorikari, R., Punie, Y.: DigComp 2.1: The digital competence framework for citizens with eight proficiency levels and examples of use,. Luxembourg 2017.

[Co17] Cohen, J., Jones, W.M., Smith, S., Calandra, B.: Makification: Towards a Framework for Leveraging the Maker Movement in Formal Education. Journal of Educational Multimedia and Hypermedia, 26(3), 217-229. Waynesville, NC USA: Association for the Advancement of Computing in Education (AACE), 2017.

[ES17] Erpenbeck, J., Sauter, W. (Eds.): Handbuch Kompetenzentwicklung im Netz: Bausteine einer neuen Lernwelt. Stuttgart: Schäffer-Poeschel Verlag,p.270, 2017.

[Fe12] Ferarri, A.: Digital Competence in Practice: An Analysis of Frameworks. JRC Technical Reports, pp.3-4, 2012.

[Fr16] Friessnig, M. H.: Development of an Educational Program Using Capabilities of (Academic) Makerspaces. International Symposium on Academic Makerspaces 2016, Cambridge, USA.

[HD16] Hochschulforum Digitalisierung: The Digital Turn – Hochschulbildung im digitalen Zeitalter (Arbeitspapier Nr. 27), Berlin, p.51, 2016.

[Jo15] Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: NMC Horizon Report: 2015 Higher Education Edition, Austin, Texas: The New Media Consortium, 2015.

[PGJ17] Papavlasopoulou, S., Giannakos, M. N., Jaccheri, L.: Empirical studies on the Maker Movement: a promising approach to learning. Entertainment Computing, 18, 57–78. 2017.

[SV16] Stifterverband: Hochschul-Bildungs-Report 2020: Hochschulbildung für die Arbeitswelt 4.0, pp.26-27, 2016.

[7] The most popular offered courses in fab labs are 3 D printing in USA and Europe [Fr16]. It combines aspects such as creativity, experimentation, open design and manufacturing etc. Therefore, it can be seen as a suitable starting point to analyse maker activities in formal academic learning.

# UnterrichtOnline.org - ein Videografieportal für alle Phasen der LehrerInnenbildung

## Interaktives und kollaboratives webbasiertes Lehren und Lernen mit Unterrichtsvideos

Robert Meyer[1] und Juliane Aulinger[2]

**Abstract:** An der UnterrichtsMitschau der LMU München entsteht derzeit ein Portal für Unterrichtsvideos, das Akteuren in der LehrerInnenaus-, -fort- und -weiterbildung sowie der pädagogischen Forschung zur Verfügung gestellt werden soll. Neben der Nutzung eines bestehenden Pools an Unterrichtvideos können auch Videos externer Projekte eingepflegt werden. Mit Hilfe einer für diesen Anwendungskontext entwickelten webbasierten Anwendung ist es dann möglich, diese Unterrichtsvideos zugriffsgeschützt in Forschung und Lehre zu nutzen.

**Keywords:** Unterrichtsaufzeichnungen, Lernplattform, Lehrerbildung

Die Videografie von Unterricht ist ein gut etabliertes Phänomen in der Schul- und Unterrichtsforschung. Bereits in den 60er Jahren des 20. Jahrhunderts entstanden an zahlreichen Universitäten sogenannte "Unterrichtsmitschauanlagen", um dem „Komplex Unterricht“ näher zu kommen [Sch70]. Auch heute wird wieder verstärkt auf Unterrichtsvideos in verschiedenen Settings zurückgegriffen, sei es in der Unterrichtsforschung [Se14] oder aber in der LehrerInnenaus-, -fort- und -weiterbildung. Großer Vorteil von Unterrichtsvideos ist dabei letztendlich immer die Dokumentation und Sicherung „flüchtiger“ Situationen im Unterricht und damit eine Möglichkeit, um der „Überkomplexität“ von Unterricht gerecht zu werden [Di16].

Seit inzwischen 50 Jahren zeichnet die Einrichtung „UnterrichtsMitschau“ der LMU München Unterricht für Forschung und Lehre in Form von authentischen Unterrichtsdokumentationen auf. Diese Unterrichtsaufzeichnungen werden zukünftig ausgewählten fachlichen NutzerInnen unter der Adresse UnterrichtOnline.org zur Verfügung gestellt. Dadurch können die Videos sowohl orts- und zeitunabhängig für E-Learning, als auch direkt in Seminaren sowie in der Forschung genutzt werden.

Für die differenzierte Bearbeitung der Unterrichtseinheiten wurde als Alleinstellungsmerkmal zu anderen Portalen eine Webanwendung geschaffen, die es den NutzerInnen ermöglicht, interaktiv und kollaborativ mit den Videos zu arbeiten. Dies wird zum einen dadurch ermöglicht, dass die Unterrichtsaufzeichnungen in mehreren

[1] LMU München, UnterrichtsMitschau, Martiusstraße 4, 80802 München, robert.meyer@lmu.de
[2] LMU München, UnterrichtsMitschau, Martiusstraße 4, 80802 München, aulinger@edu.lmu.de

doi:10.18420/delfi2019_231

Perspektiven vorliegen (z.B. Fokus auf Lehrkraft / Klasse / einzelne sprechende Schüler), welche von den unterschiedlichen Betrachtern - je nach Beobachtungsziel – interaktiv gewechselt werden können. Zum anderen erlaubt es die Software, mit Hilfe von Annotationen die Videos inhaltlich mit eigenen Anmerkungen und Metainformationen anzureichern. Die Annotationen können hier entweder nur privat oder auch öffentlich sichtbar sein. Öffentliche Annotationen erlauben eine Diskussion ähnlich wie in einem Onlineforum. Zusätzlich lassen sich die Annotationen für eine qualitative Videoanalyse verwenden: Einerseits, in dem ein vorgegebenes Kategoriensystem in der Anwendung hinterlegt wird, über welches die Beobachtungen kategorisiert und zeitlich referenziert im System gespeichert werden, andererseits, in dem auch eine dynamische Kategorienbildung mit Hilfe von #Hashtags unterstützt wird. Die Beobachtungen können dann über ein ausdifferenziertes Filtersystem direkt in der Anwendung sowie durch eine Exportfunktion mit Hilfe von Fremdsoftware ausgewertet werden. Dies ermöglicht verschiedene Lehr-Lernsettings. Beispielsweise können Studierende aufgabengeleitet Videos analysieren, ohne sich gegenseitig in ihren Beobachtungen zu beeinflussen. Vor dem zugehörigen Seminar werden die Annotationen dann veröffentlicht, damit die Beobachtungen entweder online kommentiert oder in Präsenz verglichen und diskutiert werden können.

UnterrichtOnline.org will als Videografieportal zum einen Unterrichtsvideos zugriffgeschützt für NutzerInnen in allen Phasen der Lehramtsausbildung mit einem nachgewiesenen fachlichen Interesse bereitstellen. Zusätzlich kann UnterrichtOnline.org universitätsübergreifend von Projekten genutzt werden, die selbst produzierte Videos hochladen und in den eigenen Forschungs- und Lehrkontexten verwenden möchten. Dabei sorgt die UnterrichtsMitschau für den technischen Betrieb des Portals, während die Projekte ihren Fokus auf inhaltliche und didaktische Fragen legen können.

Das Portal UnterrichOnline.org und die damit verbundenen Funktionalitäten können vor Ort von Interessierten selbst ausprobiert werden.

## Literaturverzeichnis

[Di16] Dinkelaker, J.: Datengewinnung und –formate in der videobasierten Unterrichtsforschung. In (Rauin, U.; Herrle, M.; Engartner, T., Hrsg.): Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele. Weinheim und Basel, S. 50 - 75, Beltz Juventa, 2016.

[HRE16] Herrle, M.; Rauin, U.; Engartner, T.: Videos als Ressourcen zur Generierung von Wissen über Unterrichtsrealität(en). In (Rauin, U.; Herrle, M.; Engartner, T., Hrsg.): Videoanalysen in der Unterrichtsforschung. Methodische Vorgehensweisen und Anwendungsbeispiele. Weinheim und Basel, S. 8 – 29, Beltz Juventa, 2016.

[Sch70] Schorb, A.O.: Kritische Fragen an die Unterrichtsmitschau. Schulreport 02/70, S. 3-8, 1970.

[Se14] Seidel, T.: Lehrerhandeln im Unterricht. In (Terhart, E.; Bennewitz, H.; Rothland, M., Hrsg.): Handbuch der Forschung zum Lehrerberuf (2. Auflage), Münster, S. 781 – 806, Waxmann, 2014.

Niels Pinkwart, Johannes Konert (Hrsg.): Die 17. Fachtagung Bildungstechnologien,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2019 333

# Analytics-Dashboard für Lehre, Forschung und Transfer

Jan Bernoth[1], Alexander Kiy[1] und Ulrike Lucke[1]

**Abstract:** Sämtliche Bereiche der Hochschule sind durch zunehmenden Einsatz digitaler Medien und Systeme gekennzeichnet. Dabei fallen in verschiedenen Systemen Daten an, die visuell und semantisch in Form von Analytics-Dashboards aufbereitet von den Nutzenden zur Optimierung von Kernprozessen aus Forschung und Lehre verwendet werden können. Während Learning Analytics lediglich auf den Teilbereich Lehre und Studium abzielt, werden in diesem Posterbeitrag auch Prozesse für Forschung und Transfer betrachtet. Darüber hinaus werden Ideen für ein hochschulweit nutzbares Analytics-Dashboard sowie entwickelte Anwendungsszenarien präsentiert.

**Keywords:** Analytics, Learning Analytics, Dashboard, Data Visualisation.

## 1 Motivation

Durch die zunehmende Verwendung digitaler Medien und Systeme werden sogenannte Analytics-Dashboards [SRV+17] zu einem zentralen Element der Visualisierung von Daten, die durch Nutzende generiert wurden. Im Hochschulbereich wird Analytics meist im Bereich von Studium und Lehre betrachtet und unter dem Begriff *Learning Analytics* (LA) subsumiert. Der Forschungsbereich hingegen wird mit dem Analytics-Gebiet *Bibliometrics* & *Scientometrics* abgedeckt [Coo12]. Analytics-Komponenten oder Dashboards mit Fokus auf Forschung, Transfer oder Überschneidungsbereiche (zum Beispiel Lehre & Transfer) scheinen bisher unterrepräsentiert zu sein. Mit dem Ziel ein Analytics-Dashboard für ein zentrales Hochschulportal zu entwickeln, das gleichermaßen für Lehre, Forschung und Transfer genutzt werden kann, wurde ein Workshop durchgeführt, in dem Studierende, wissenschaftliche und technische Mitarbeitende gemeinsam Ideen für Analytics-Komponenten skizzieren.

Im Folgenden werden die gesammelten Ideen vorgestellt und mittels des Learning Analytics Life Cycle von Khalil und Ebner [KE15] klassifiziert. Dieser unterteilt sich in vier Phasen: Learning Environment (*Stakeholder*), Big Data (*Datenquellen*), Analytics (*Techniken*) und Act (*Ziele*). Durch die Klassifizierung werden die einzelnen Ideen gebündelt und Visualisierungstechniken oder Datenquellen können ermittelt werden, die für mehrere Analytics-Komponenten wichtig sind. Die Ideen werden anschließend hinsichtlich ihrer zu erwartenden Nützlichkeit gewichtet. Aus der Häufigkeit der Nennung, der Beliebtheit und einer Machbarkeitsanalyse ergibt sich ein mehrdimensionales Anforderungsgefüge für die Anfertigung eines Prototyps.

[1] Universität Potsdam, Institut für Informatik & Computational Science, August-Bebel-Str. 89, 14482 Potsdam, vorname.nachname@uni-potsdam.de

doi:10.18420/delfi2019_354

## 2 Ergebnisse

An dem Workshop nahmen 8 Personen teil, die gemeinsam 31 Komponenten konzipierten. Die Komponenten sind für Studierende, Lehrende, Forschende und technisch Mitarbeitende konzipiert worden und zeigen zum Beispiel Personeninteraktionen in einer Gruppenarbeit, Flaschenhals-Probleme im Zusammenspiel verschiedener Systeme oder eine Auslastungsübersicht der Mensa. Die vollständigen Ergebnisse und das entstandene Poster können auch im öffentlichen Repository eingesehen werden.[2] Zu den einzelnen Ideen wurden die Meinungsbilder zwei verschiedener Gruppen gegenübergestellt. Die erste Gruppe bestand aus E-Learning-affinen Teilnehmenden (N=12), die zweite Gruppe aus Studierenden unterschiedlicher Studiengänge (N=11). Die zwei disjunkten Gruppen wurden gewählt, um mögliche Verzerrungseffekte durch thematische Nähe oder Vorwissen zu erkennen. In beiden Gruppen beliebte Visualisierungskategorien waren: Visualisierungen des Zustands der zentralen Dienste und Übersichten über den Arbeitseinsatz von Tutoren.

Für die Kategorisierung der Komponenten wurden die Dimensionen des LA Life Cycle nach [KE15] verwendet, da sich noch keine Taxonomie in Learning Analytics durchgesetzt hat. Durch die zusätzlichen Bereiche Forschung und Transfer werden keine neuen Dimensionen benötigt, hingegen müssen für die neu entstandenen Anforderungen innerhalb der Dimensionen neue Kategorien entstehen. Zum Beispiel fehlen in der Dimension *Datenquellen* bislang Zeitmanagementdaten, Evaluationsdaten, technische Systemdaten und hochschulbezogene Daten. Mit einer aktualisierten, vervollständigten und trennscharfen Kategorisierung könnte eine integrierte Datenbank mit Visualisierungsmöglichkeiten geschaffen werden, die sich für ein umfassendes Analytics-Dashboard für Lehre, Forschung und Transfer eignet. Dies würde die Einstiegshürden für Hochschulen verringern und den Datenaustausch zwischen Hochschulen fördern.

Literaturverzeichnis

[Coo12] Cooper, A.: CETIS Analytics Series: A Brief History of Analytics. vol. 1, no. 9. University of Bolton, doi: 10.1.1.269.6470.

[KE15] Khalil, M.; Ebner, M.: Learning Analytics: Principles and Constraints. Montreal, Quebec, Canada, Association for the Advancement of Computing in Education (AACE), S. 1789–1799, 2015.

[SRV+17] Schwendimann, B. A.; Rodríguez-Triana, M. J.; Vozniuk, A.; Prieto, L. P.; Boroujeni, M. S.; Holzer, A.; Gillet, D.; Dillenbourg, P.: Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. IEEE Transactions on Learning Technologies, vol. 10, no. 1, S. 30–41, 2017.

[2] https://doi.org/10.5281/zenodo.3253190

# Welche inhaltliche Unterstützung ist für Studierende beim Programmieren in einem Remote-Labor hilfreich?

Anja Hawlitschek[1], André Dietrich[2] und Sebastian Zug[3]

**Abstract:** Der webbasierte Zugriff auf reale Experimente gibt dem Lernenden die Möglichkeit, Aufgaben zu beliebigen Zeitpunkten zu realisieren. Gleichzeitig muss das System aber auch die fehlende unmittelbare Interaktion mit einem Lehrenden kompensieren. Folglich sollte das Aufgabendesign und die Umgebung den individuellen Unterstützungsbedarf adressieren. Das Paper evaluiert hierzu den Einfluss unterschiedliche Detailgrade einer Aufgabenstellung und Unterstützungsmethodik.

**Keywords:** Remote-Labor, Nutzerspezifische Unterstützung

## 1 Einleitung

Eine didaktische Herausforderung bei Remote-Laboren besteht darin, Studierenden die notwendige inhaltliche Anleitung und Unterstützung zu bieten. Mit der fehlenden Möglichkeit der synchronen Kommunikation mit Lehrenden können inhaltliche Probleme nicht zeitnah besprochen werden. Die Unterstützung des Lernenden (Anleitungen, Hilfestellungen) sollte deshalb beim Design der Lerninhalte mitgedacht werden, um Frustration und Dropout zu vermeiden. Zu detaillierte Ausführungen können allerdings dazu führen, dass Lernende sich weniger kognitiv engagieren und in geringem Maße eigene Erklärungen für kognitive Herausforderungen generieren [RN13]. Zudem sinkt ggf. die intrinsische Motivation [RJ06]. Um diese didaktische Herausforderung im Kontext des Lernens zu untersuchen, haben wir in zwei experimentellen Studien[4] zwei Formen von Anleitung für eine Informatikübung implementiert und die Wirkungen auf die Motivation und den Programmier- bzw. Lernerfolg untersucht.

## 2 Studien

Die Studierenden müssen in den Selbstlernphasen unterschiedliche Aufgaben mit kleinen Robotersystemen im Remote-Labor lösen, die auf Webseiten im System beschrieben werden. In Studie 1 (N = 58) integrierten wir für die Experimentalgruppe (N = 29) im Rahmen einer Aufgabe zusätzlich sogenannte Prompts, d.h. Hinweise welche die

[1] Hochschule Magdeburg-Stendal, Breitscheidstraße 2, 39114 Magdeburg, anja.hawlitschek@hs-magdeburg.de
[2] Otto-von-Guericke Universität, Universitätsplatz 2, 39106 Magdeburg, andre.dietrich@ovgu.de
[3] TU Bergakademie Freiberg, Bernhard-von-Cotta-Straße 2, 09596 Freiberg, sebastian.zug@informatik.tu-freiberg.de
[4] Im Rahmen des Projekts "Industrial-eLab", gefördert durch das BMBF (Fördernummer 16DHL1034), siehe Projektwebseite http://www.elab.ovgu.de

doi:10.18420/delfi2019_342

Lernenden auf relevante Fakten, Prozeduren und Ansätze aufmerksam machen sollen [Th09]. So sollen typische Programmierfehler vermieden oder auf effiziente Programmierwege aufmerksam gemacht werden. In der zweiten Studie im Rahmen einer 90 minütigen Lehrveranstaltung (N = 27) fokussierten wir auf den Programmierprozess. Schwer verständliche Compiler-Fehler-Meldungen werden als ein Grund für hohe Dropout-Raten und geringe Erfolge in Programmierkursen angesehen [FZJ04]. Aus diesem Grund implementierten wir ausführlichere Compiler-Fehler-Meldungen, die der Experimentalgruppe (N = 13) zur Verfügung standen.

## 3 Diskussion der Ergebnisse

ANOVAs ergeben, dass Prompts in Studie 1 keine signifikanten Wirkungen auf die Performanz beim Programmieren ($F(1,54) = 0.18$, $p = .67$, $\eta^2 = .00$) und die Abschlussnote ($F(1,26) = .36$, $p = .55$, $\eta^2 = .01$) sowie einen negativen Effekt auf die selbsteingeschätzte intrinsische Motivation ($F(1,37) = 4.22$, $p = .04$, $\eta = .11$) haben. Dieses Ergebnis deutet an, dass mehr Anleitung nicht immer zu mehr Lernerfolg führt, sondern auch eine motivationsmindernde Einschränkung der Lernenden, im Versuch selbsttätig Probleme zu lösen, bedeuten kann. Die ausführlicheren Compiler-Fehler-Meldungen in Studie 2 dagegen führten zu signifikant besseren Programmiererfolgen ($F(1,26) = 6.68$, $p = .02$, $\eta^2 = .21$), ohne die intrinsische Motivation zu beeinträchtigen ($F(1,26) = 1.45$, $p = .24$, $\eta^2 = .05$), und scheinen daher ein für die Praxis geeignetes Mittel für die Unterstützung von Lernenden in Programmierprozessen – auch in Remote Laboren – zu sein. Die Ergebnisse beider Studien zeigen einmal mehr, wie herausfordernd es ist, das richtige Maß und die richtige Form von instruktionaler Anleitung für Studierende zu finden.

Literaturverzeichnis

[FZJ04] Flowers, T.; Carver, C. A. & Jackson, J.: Empowering students and building confidence in novice programmers through gauntlet. 34th Annual Frontiers in Education. FIE, 2004.

[RJ06] Reeve, J. & Jang, H.: What teachers say and do to support students' autonomy during a learning activity. Journal of Educational Psychology, 98(1), 209-218, 2006.

[RN13] Richey, E. & Nokes-Malach, T. J.: How much is too much? Learning and motivation effects of adding instructional explanations to worked examples. Learning and Instruction, 25, 104-124, 2013.

[Th09] Thillmann, H.; Künsting, J.; Wirth, J. & Leutner, D.: Is it merely a question of ‘what’ to prompt or also ‘when’ to prompt? - The role of point of presentation time of prompts in self-regulated learning. Zeitschrift für Pädagogische Psychologie, 23(2), 105–115, 2009.

# Verzeichnis der Autorinnen und Autoren


*Allert, Heidrun* ..... 229
*Ashimi, Yasmin* ..... 271
*Aulinger, Juliane* ..... 331
*Bade, Korinna* ..... 285
*Bartels, Mareike* ..... 325
*Bau, Beatrix* ..... 163
*Bauer, Mathias* ..... 163
*Berens, Florian* ..... 297
*Bergner, Nadine* ..... 319
*Bernoth, Jan* ..... 333
*Biedermann, Daniel* ..... 57
*Blasberg, Alexander* ..... 327
*Braeschke, Lucas* ..... 73
*Brandenburger, Bonny* ..... 329
*Braun, Iris* ..... 73
*Busse, Julian* ..... 67
*Böhnke, Stephanie* ..... 285
*Darabi, Sima* ..... 299
*Dietrich, André* ..... 335
*Drachsler, Hendrik* ..... 57, 175
*Durski, Sara* ..... 271
*Ehlenz, Matthias* ..... 287, 319
*Ehlenz, Matthias* ..... 303
*Eichhorn, Michael* ..... 85
*Engel, Mailyn* ..... 285
*Fortenbacher, Albrecht* .....193, 199
*Frey, Lisa* .....271
*Froitzheim, Manuel* .....217, 313
*Funk, Johannes* .....145
*Garmann, Robert* .....103
*Geßner, Hendrik* .....187
*Gips, Carsten* .....127
*Goedicke, Michael* .....91, 327
*Hanck, Christoph* .....327
*Hara, Tenshi* .....73
*Hawlitschek, Anja* .....335
*Helbig, René* .....193
*Hobert, Sebastian* .....297
*Hobert, Sebastian* .....259
*Hoffmann, Andreas* .....217
*Hoffmann, Michael* .....157
*Huckfeldt, Finja* .....325
*Hübl, Reinhold* .....121
*John, Elias* .....235
*Judel, Sven* .....307
*Kapp, Felix* .....73
*Keus, Alexandra* .....15
*Kiy, Alexander* .....333
*Kiy, Alexander* .....187
*Klingauf, Anna* .....145

*Konert, Johannes* ..... 293
*Korth, Susanne* ..... 21
*Krömker, Heidi* ..... 163
*König, Matthias* ..... 127
*Küntzer, Lucas* ..... 157
*Lange, Aline* ..... 67
*Larisch, Kathrin* ..... 315
*Lechtenbörger, Jens* ..... 205
*Lefarth, Justin* ..... 91
*Leonhardt, Thiemo* ..... 287, 303, 319
*Lingnau, Andreas* ..... 253
*Lorenz, Uwe* ..... 321
*Lucke, Ulrike* ..... 289, 305, 333
*Lutz, Matthias* ..... 127
*Lübcke, Eileen* ..... 325
*Lüüs, Angela* ..... 145
*Magdowski, Mathias* ..... 295
*Maier-Rothe, Jörg* ..... 277
*Massing, Till* ..... 327
*Massler, Ute* ..... 271
*Meyer, Robert* ..... 331
*Moebert, Tobias* ..... 289
*Moeller, Korbinian* ..... 193
*Mrech, Heike* ..... 285
*Müller, Eric* ..... 301
*Müller, Wolfgang* ..... 271
*Naoufal, Ala* ..... 293
*Naoufal, Hicham* ..... 293
*Niemeyer, Jana* ..... 85
*Ninaus, Manuel* ..... 193
*Noichl, Svenja* ..... 15, 21, 309
*Otto, Benjamin* ..... 91, 327
*Palsbröker, Patrick* ..... 127
*Patzer, Yasmin* ..... 311
*Pfeiffer, Anke* ..... 323
*Pfennig, Anja* ..... 277
*Pinkwart, Niels* ..... 311
*Podgayetskaya, Tatyana* ..... 121
*Reckmann, Natalie* ..... 327
*Reimann, Gabi* ..... 325
*Reinstorf, Tatjana* ..... 321
*Rensing, Christoph* ..... 45
*Riazy, Shirin* ..... 199, 223
*Richter, Christoph* ..... 229
*Richter, Dirk* ..... 133
*Richter, Eric* ..... 133
*Rieger, Christian Moritz* ..... 33
*Rittberger, Marc* ..... 175
*Rock, Georg* ..... 157
*Rustemeier, Linda* ..... 27

*Röpke, René* .. 287, 307, 315, 317, 319
*Sauer, Marc* ..... 217
*Scheidat, Tobias* ..... 285
*Schmidt, Ludger* ..... 145
*Schmiemann, Philipp* ..... 91
*Schneider, Jan* ..... 175
*Scholz, Sandra* ..... 285
*Schramm, Thilo* ..... 91
*Schroeder, Ulrik*
.. 15, 21, 287, 307, 309, 315, 317, 319
*Schröder, Nadine* ..... 291
*Schuhen, Michael* ..... 313
*Schuldt, Jacqueline* ..... 163
*Schumann, Matthias* ..... 67
*Schwarzbach, Marguerite* ..... 271
*Schwarzer, Franz* ..... 301
*Schöbel, Sven* ..... 315
*Seidel, Niels* ..... 33
*Seiffert, Sven-Bjarne* ..... 253
*Simbeck, Katharina* ..... 199, 223
*Steuer, Tim* ..... 45
*Striewe, Michael* ..... 91
*Striewe, Michael* ..... 115
*Söbke, Heinrich* ..... 247
*Tillmann, Alexander* ..... 85
*Uckelmann, Dieter* .....323
*van den Berk, Ivo* .....325
*Webers, Marie Jeanne* .....163
*Weicker, Karsten* .....301
*Weigel, Miriam* .....121
*Weise, Matthias* .....305
*Westburg, Anika* .....283
*Wiepke, Axel* .....133
*Wollny, Sebastian* .....175
*Yun, Haeseon* .....193, 199
*Zaric, Nadja* .....307
*Zender, Raphael* .....133, 305
*Zug, Sebastian* .....335