

GESELLSCHAFT  
FÜR INFORMATIK





Raphael Zender, Dirk Ifenthaler,  
Thiemo Leonhardt, Clara Schumacher (Hrsg.)

**DELFI 2020 – Die 18. Fachtagung  
Bildungstechnologien  
der Gesellschaft für Informatik e.V.**

**14.-18. September 2020  
Online**

Gesellschaft für Informatik e.V. (GI)

## **Lecture Notes in Informatics (LNI) - Proceedings**

Series of the Gesellschaft für Informatik (GI)

Volume P-308

ISBN 978-3-88579-702-9

ISSN 1617-5468

### **Volume Editors**

Dr.-Ing. Raphael Zender

Universität Potsdam

Am Neuen Palais 10, 14469 Potsdam, Deutschland

raphael.zender@uni-potsdam.de

Prof. Dr. Dirk Ifenthaler

Dr. rer. nat. Thiemo Leonhardt

Dr. Clara Schumacher

### **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

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**

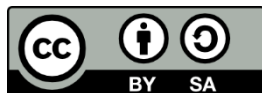
Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

### **Seminars**

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

© Gesellschaft für Informatik, Bonn 2020

printed by Köllen Druck+Verlag GmbH, Bonn



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



## Educational Realities

Wie in fast allen gesellschaftlichen Bereichen findet auch die Nutzung von Technologien im Bildungskontext nicht einheitlich Einzug in den Alltag, sondern in unterschiedlichen Geschwindigkeiten. Vergleicht man dies mit einer Autobahn, finden sich auf der linken Spur vor allem Visionen und Fallstudien in besonders experimentierfreudigen Lehr- und Lernkontexten – oft vor allem in der beruflichen Bildung einzelner Institutionen und der Weiterbildung in Technologie-Startups. Auf der Mittelspur holen punktuell gerade Hochschulen im Rahmen von Forschungsprojekten zunehmend auf, auch wenn deren technologische Neuerungen es erst spät in den Regelbetrieb der Bildungseinrichtung schaffen. Dafür finden wir hier viele fundierte Untersuchungen über den technologischen Nutzen. Auf der rechten und überaus breiten Fahrspur müssen wir den tatsächlichen Bildungsalltag in der Mehrzahl der beruflichen und (hoch)schulischen Bildungskontexte verorten.

Den unterschiedlichen Geschwindigkeiten dieser "Bildungsautobahn" und deren Implikationen widmet sich die 18. Fachtagung Bildungstechnologien (DELFI 2020) der GI-Fachgruppe Bildungstechnologien. Das diesjährige Tagungsmotto lautet daher *Educational Realities*. Durch das Motto wird die tatsächliche Verortung technologiegestützter Bildungsprozesse zwischen einem oft ernüchterndem Bildungsalltag und teils realitätsfernen Bildungsvisionen thematisiert. Somit eröffnet die DELFI ein interdisziplinäres Spannungsfeld von alltäglichen Herausforderungen durch Bildungstechnologien über empirisch evidente Reflektionen zu deren Einsatz bis hin zu realitätsüberschreitenden Trends, insbesondere im Bereich Virtual Reality, Augmented Reality und Learning Analytics.

Die DELFI 2020 findet vom 14. bis 18. September 2020 Corona-bedingt erstmals vollständig online statt. Durch die erstmalige Austragung der DELFI in Zusammenarbeit mit der 15th annual European Conference on Technology-Enhanced Learning (EC-TEL) zeichnet sich die DELFI in 2020 auch durch eine deutlich internationalere Perspektive als in den Vorjahren aus.

Forschende und Praktizierende aus allen Fachdisziplinen wurden eingeladen, ihre Beiträge und Ergebnisse einzureichen und dem Fachpublikum zur Diskussion vorzustellen. Beiträge konnten auf Deutsch oder Englisch verfasst werden, wodurch sich in diesem Tagungsband sowohl deutschsprachige als auch englischsprachige Texte wiederfinden.

Die eingereichten wissenschaftlichen Langbeiträge wurden im Doppelblind-Verfahren von je mindestens drei Mitgliedern des Programmkomitees begutachtet. Bei einem Programmkomitee-Treffen wurde dann über die Annahme und Ablehnung von Beiträgen entschieden. Von 36 eingereichten wissenschaftlichen Langbeiträgen wurden 13 angenommen. Dies entspricht eine Annahmequote von ca. 36 %. Die angenommenen Langbeiträge finden sich im vorliegenden Tagungsband, ergänzt durch weitere Beitragsformen – 15 Kurz-, 13 Praxis-, 12 Poster-, 7 Demobeiträge sowie 2 Beiträge zu den DELFI-Keynotes.

Wir freuen uns ganz besonders, dass wir wieder zwei renommierte Keynotes gewinnen konnten. Prof. Dr. Sabine Seufert (University of St. Gallen) wird mit uns einen Schritt zurück ("Back to the roots") wagen, um eine fundiertere digitale Transformation zu errei-

chen. Prof. Dr. Samuel Greiff (University of Luxembourg) wird aktuelle Entwicklungsfelder, Chancen und Herausforderungen bei der Bewertung von Bildungstechnologien im Hinblick auf die vielgenannte "Digitale Revolution" identifizieren und diskutieren.

Die DELFI 2020 wird zudem von sechs Workshops begleitet, die sich konzentriert einem tagungsrelevanten Teilgebiet widmen und teils in Kooperation mit EC-TEL-Workshops stattfinden. Zudem findet auch 2020 wieder ein Doktorandenkolloquium zur Förderung des wissenschaftlichen Nachwuchses im Bereich Bildungstechnologien statt.

Zu guter Letzt möchten wir im Namen der Chairs und des Organisations-Teams all jenen unseren Dank aussprechen, die zum Gelingen dieser Veranstaltung beigetragen haben. Erst durch die vielen Beitragenden, die die Ergebnisse ihrer Arbeit als Tagungsbeiträge aufbereitet haben, wurde die DELFI 2020 inkl. ihrer Workshops möglich! Ebenso gilt unser Dank den Mitgliedern des Programmkomitees, die diese Einreichungen aus ihrer jeweiligen Perspektive geprüft haben und auf der Sitzung des Programmkomitees eine nicht immer leichte Auswahl treffen mussten. Für die "lokale" Organisation möchten wir Marco Kalz (PH Heidelberg) und seinem Team danken, welches flexibel und kreativ den Wandel der DELFI und EC-TEL zur Online-Tagung vollzogen hat. Dieser Dank schließt auch die vielen Helfer ein, ohne die ein reibungsloses und angenehmes Konferenzerlebnis undenkbar wäre – erst recht in der aktuellen Online-Form.

Wir wünschen uns allen eine erkenntnisreiche und schöne DELFI und EC-TEL mit vielen interessanten Beiträgen und mit inspirierenden Begegnungen.

Potsdam und Mannheim im September 2020

Raphael Zender, Universität Potsdam  
Dirk Ifenthaler, Universität Mannheim

## Tagungsleitung

Raphael Zender, Universität Potsdam (Chair)  
Dirk Ifenthaler, Universität Mannheim (Chair)  
Thiemo Leonhardt, Technische Universität Dresden (Poster- und Demo-Chair)  
Clara Schumacher, Humboldt-Universität zu Berlin (Workshop Chair)

## Programmkomitee

Timo Ahlers  
Universität Potsdam

Daniel Bodemer  
Universität Duisburg-Essen

Andreas Breiter  
Universität Bremen

Ilona Buchem  
Beuth Hochschule für Technik Berlin

Mohamed Amine Chatti  
Universität Duisburg Essen

Mario Donick  
vFlyteAir Simulations

Ralf Dörner  
Hochschule RheinMain

Hendrik Drachslers  
DIPF

Jens Drummer  
Sächsisches Bildungsinstitut

Martin Ebner  
Technische Universität Graz

Wolfgang Effelsberg  
Universität Mannheim

Albrecht Fortenbacher  
HTW Berlin

Paul Grimm  
Hochschule Fulda

Josef Guggemos  
Universität St. Gallen

Jörg Haake  
Fernuniversität in Hagen

Sebastian Habig  
Universität Duisburg-Essen

Andreas Harrer  
Fachhochschule Dortmund

Peter Henning  
Hochschule Karlsruhe

Ulrich Hoppe  
Universität Duisburg-Essen

Marco Kalz  
Pädagogische Hochschule Heidelberg

Paul-Thomas Kandzia  
Duale Hochschule Baden-Württemberg

Michael Kerres  
Universität Duisburg-Essen

Andrea Kienle  
Fachhochschule Dortmund

Ralf Klamma  
RWTH Aachen University

Ralph Koelle  
Universität Hildesheim

Johannes Konert  
Beuth Hochschule für Technik Berlin

Bernd Krämer  
Fernuniversität in Hagen

Maike Krannich  
Universität Zürich

Detlef Krömker  
Goethe Universität Frankfurt am Main

Roland Küstermann  
Duale Hochschule Baden-Württemberg

Nguyen-Thanh Le  
Humboldt-Universität zu Berlin

Carsten Lecon  
Hochschule Aalen

Andreas Lingnau  
Hochschule Ruhr West

Jörn Loviscach  
Fachhochschule Bielefeld

Ulrike Lucke  
Universität Potsdam

Vlatko Lukarov  
RWTH Aachen University

Dana-Kristin Mah  
Stifterverband

Alke Martens  
Universität Rostock

Christoph Meier  
Universität St. Gallen

Christoph Meinel  
Hasso-Plattner-Institut

Agathe Merceron  
Beuth Hochschule für Technik Berlin

Helmut Niegemann  
Universität des Saarlandes

Thies Pfeiffer  
Hochschule Emden/Leer

Niels Pinkwart  
Humboldt-Universität zu Berlin

Hans Pongratz  
Technische Universität München

Anja Richert  
Technische Hochschule Köln

Holger Rohland  
Technische Universität Dresden

Guido Rößling  
Technische Universität Darmstadt

Nikol Rummel  
Ruhr-Universität Bochum

Pia Sander  
Universität Duisburg Essen

Daniel Schiffner  
DIPF

Jan Schneider  
DIPF

Mareike Schoop  
Universität Hohenheim

Ulrik Schroeder  
RWTH Aachen University

Sandra Schulz  
Humboldt-Universität zu Berlin

Andreas Schwill  
Universität Potsdam

Niels Seidel  
Fernuniversität in Hagen

Heinrich Söbke  
Bauhaus-Universität Weimar

Pia Spangenberg  
Technische Universität Berlin

Sven Strickroth  
Universität Potsdam

Michael Striwe  
Universität Duisburg-Essen

Alexander Tillmann  
Goethe Universität Frankfurt am Main

Stephan Trahasch  
Hochschule Offenburg

Dimitra Tsovaltzi  
DFKI Saarbrücken

Markus von der Heyde  
vdH-IT

Jane Yau  
Universität Mannheim

## **Sub-Reviewer**

Julian Dehne  
Universität Potsdam

Michael Eichhorn  
Goethe Universität Frankfurt am Main

Stefanie Elbeshausen  
Universität Hildesheim

Melanie Erkens  
Universität Duisburg-Essen

Carolin Hainke  
Hochschule Emden/Leer

Jule Marleen Krüger  
Universität Duisburg-Essen

Leonard Meyer  
Universität Bielfeld

Angela Rizzo  
Goethe Universität Frankfurt am Main

Lenka Schnaubert  
Universität Duisburg-Essen



# Inhaltsverzeichnis

## Keynotes

**Sabine Seufert**

*Current and Emerging Educational Realities - Shaping the Digital Transformation?* ..... 19

**Samuel Greiff**

*Technology-Based Assessment in 21st Century Education* ..... 21

## Forschungs- und Praxisbeiträge

**Johannes Funk, Ludger Schmidt**

*Evaluation einer Augmented-Reality-Anleitung für eine komplexe Montageaufgabe* ..... 25

**Niklas Osmers, Michael Prilla**

*Potential vs. Practice: Challenges for the Implementation of Augmented Reality for Learning and Training in Practice* ..... 37

**Jens Hofmann**

*LEAP: Learnings beim Einsatz von Augmented Reality in der praktischen Berufsausbildung* ..... 49

**Leonard Meyer, Thies Pfeiffer**

*Comparing Virtual Reality and Screen-based Training Simulations in Terms of Learning and Recalling Declarative Knowledge* ..... 55

**Carsten Fussen, Veronika Hanesova**

*Pre-Testing der Lerneffektivität von 2D-to-3D-Didactics in immersiven VR-Umgebungen* ..... 67

**Tobias Leßner**

*Demokratische Schule 360°: Potenziale Virtueller Realität für LehrerInnenbildung und Ethnografie* ..... 73

**Christian Plotzky, Ulrike Lindwedel-Reime, Alexander Bejan, Peter König, Christophe Kunze**

*Virtual Reality in Health Care Education: A Study about the Effects of Presence on Acceptance and Knowledge Improvement among Health Care Students* ..... 79

**Nikolaj Troels Graf von Malotky, Alke Martens**

*DigiCare - Intelligentes Tutoring für Gesundheitsmanagement* ..... 91

**Thiemo Leonhardt, Gregor Damnik, Nadine Bergner**

*Touch-Aktionen beim digitalen Lernen: Steigerung der Performance des visuellen Gedächtnisses durch aktive Reizverarbeitung* ..... 97

<b>Jonas Blattgerste, Kristina Luksch, Carmen Lewa, Martina Kunzendorf, Nicola H. Bauer, Annette Bernloehr, Matthias Joswig, Thorsten Schäfer, Thies Pfeiffer</b>	
<i>Project Heb@AR: Exploring handheld Augmented Reality training to supplement academic midwifery education.....</i>	103
<b>Michael Striewe, Florian Trauten, Carolin Eitemüller</b>	
<i>Aufgaben mit automatischem Feedback zu chemischen Atom-Orbitalmodellen.....</i>	109
<b>Florian Horn, Daniel Schiffner, Patrick Sacher, Thorsten Gattinger</b>	
<i>Usability design and evaluation for a formative assessment feedback .....</i>	121
<b>Orlando Brix, Susann Schober, Leena Steinke, Sven Strickroth</b>	
<i>Biblisches Hebräisch digital lehren und lernen: Praxiserfahrungen beim Aufbau eines Blended-Learning-Formats mit Moodle.....</i>	127
<b>Sven Judel, Timo Bergerbusch, Ulrik Schroeder</b>	
<i>Automatisierte Generierung von Automaten und automatenbasierten Aufgaben ...</i>	133
<b>Joerg M. Haake, Niels Seidel, Heike Karolyi, Lihong Ma</b>	
<i>Self-Assessment mit High-Information Feedback .....</i>	145
<b>Dirk von Suchodoletz, Sven Slotosch, Christian Rößler, Steffen Ritter</b>	
<i>Umsetzung digitaler Prüfungen mit bwLehrpool an der Universität Freiburg.....</i>	151
<b>Saba Mateen, Sarah Voß-Nakkour, Linda Rustemeier</b>	
<i>Eine Studie zur Qualitätsbeurteilung von automatisierten Testwerkzeugen zur Prüfung auf Barrierefreiheit .....</i>	157
<b>Tornike Giorgashvili, Sarah Voß-Nakkour</b>	
<i>Nutzerzentrierter Ansatz zur Behebung von digitalen Barrieren .....</i>	169
<b>Arlett Semm, Ulrike Spierling</b>	
<i>Ermittlung von Usability-Kriterien für Jugendliche mit kognitiven Beeinträchtigungen im Bereich E-Learning .....</i>	181
<b>Natalie Kiesler</b>	
<i>Kompetenzmodellierung für die grundlegende Programmierausbildung - Eine kritische Diskussion zu Vorzügen und Anwendbarkeit der Anderson Krathwohl Taxonomie im Vergleich zum Kompetenzmodell der GI.....</i>	187
<b>Tobias Falke</b>	
<i>Erwerb von Digitalen Kompetenzen für die Lebens- und Arbeitswelt .....</i>	193
<b>Franziska Günther, Jörg Neumann, Robert Lorenz, Annegret Umlauf</b>	
<i>Medienkompetenz fördern und Medieneinsatz befördern – Didaktische Überlegungen bei der Entwicklung einer mobilen Anwendung für das betriebliche Ausbildungspersonal in KKUs.....</i>	199
<b>Annett Thüring, Kathrin Jäger</b>	
<i>Objektorientierte Programmierung - Kompetenzerwerb im Mastery Model des Inverted Classroom.....</i>	205



<b>Heide Faeskorn-Woyke, Birgit Bertelsmeier, Jan Strohschein</b> <i>A Decision Tree Approach for the Classification of Mistakes of Students Learning SQL, a case study about SELECT statements.....</i>	211
<b>Frank Höppner</b> <i>Taking benefit from fellow students code without copying off -- making better use of students collective work.....</i>	217
<b>Richard Lenz, David Haller, Andreas Wahl</b> <i>Maßnahmen zur Verbesserung der Effektivität von Blended Learning bei Datenbank-Vorlesungen .....</i>	223
<b>Patrick Stalljohann, Maik Merten</b> <i>VITMaze – Die Java Coding-Challenge für Verwaltungsinformatiker .....</i>	229
<b>Sebastian Serth, Ralf Teusner, Christoph Meinel</b> <i>Digitale Arbeitsblätter mit interaktiven Programmieraufgaben im Informatik-Unterricht .....</i>	235
<b>Bernadette Spieler, Carina Girvan</b> <i>Das PECC-Framework: Gender-Sensibilität und spielerische Programmierung in der informatischen Grundbildung.....</i>	247
<b>Markus Dahm, Jennifer Rose, Marius Köhler</b> <i>Programmier-Praktikum für Erstsemester – Erfahrungen aus mehreren Iterationen.....</i>	259
<b>Sylvio Rüdian, Joachim Quandt, Kathrin Hahn, Niels Pinkwart</b> <i>Automatic Feedback for Open Writing Tasks: Is this text appropriate for this lecture? .....</i>	265
<b>Sylvio Rüdian, Alexander Heuts, Niels Pinkwart</b> <i>Educational Text Summarizer: Which sentences are worth asking for?.....</i>	277
<b>Sylvio Rüdian, Niels Pinkwart</b> <i>Is the context-based Word2Vec representation useful to determine Question Words for Generators? .....</i>	289
<b>Marcel Lehmann, Timon Bauer, Stefanie Kersten, Sven Hofmann, Silvia Schöneburg-Lehnert</b> <i>Was bedeutet „Digitaler Unterricht“? .....</i>	295
<b>Shirin Riazzy, Katharina Simbeck, Marco Träger, Robert Wöstenfeld</b> <i>Mobile First: Trends in Virtual Learning Environments .....</i>	301
<b>Florian Funke, Sven Hofmann</b> <i>Digitalisierung von Analyse- und Auswertungsstrukturen im Kontext schulischer Wettbewerbsszenarien.....</i>	307
<b>Helena Barbas, Esther Bender, Fabian Hamann, Daniel Sitzmann, Marcus Soll</b> <i>Fertigkeiten erkennen und Interesse für Informatik wecken: Der Informatiktest von MINTFIT Hamburg .....</i>	313

<b>Julian Dehne, Sven Strickroth, Ulrike Lucke</b> <i>FL-Trail: Gruppenbasiertes forschendes Lernen digital unterstützen.....</i>	319
<b>Matthias Ehlenz, Birte Heinemann, Thimeo Leonhardt, René Röpke, Vlatko Lukarov, Ulrik Schroeder</b> <i>Eine forschungspraktische Perspektive auf xAPI-Registries .....</i>	331
<b>Nadine Schröder, Peter Pfänder</b> <i>Nutzung von GitHub für Open Educational Resources.....</i>	337
<b>Linda Blömer, Christin Voigt, Uwe Hoppe</b> <i>Corona-Pandemie als Treiber digitaler Hochschullehre.....</i>	343
<b>Poster</b>	
<b>Andre Greubel, Theresa Rudolph, Martin Hennecke</b> <i>VeraSQL: An educational client for relational databases .....</i>	351
<b>Mario Wolf, Heinrich Söbke, Florian Wehking, Stefanie Hörnlein</b> <i>360-degree Models in Environmental Engineering Education: an Explorative Case Study.....</i>	353
<b>Carolin Hainke, Thies Pfeiffer</b> <i>Adapting virtual trainings of applied skills to cognitive processes in medical and health care education within the DiViFaG project .....</i>	355
<b>Arno Wilhelm-Weidner</b> <i>Selbsteinschätzung leicht gemacht - Studienergebnisse aus der Theoretischen Informatik.....</i>	357
<b>Yusra Tehreem, Thies Pfeiffer</b> <i>Immersive Virtual Reality Training for the Operation of Chemical Reactors .....</i>	359
<b>Sebastian Claus, Niels Pinkwart</b> <i>LAYA will in die Schule – Eine Anforderungsanalyse für den inklusiven, kollaborativen Einsatz einer Lernsoftware in der Sekundarstufe .....</i>	361
<b>Sara Durski, Ute Massler, Wolfgang Müller</b> <i>Digitales Lesetheater zur Förderung der Leseflüssigkeit .....</i>	363
<b>Marvin Kastner, Janna Franzkeit, Anna Lainé</b> <i>Teaching Machine Learning and Data Literacy to Students of Logistics using Jupyter Notebooks .....</i>	365
<b>Kensuke Akao</b> <i>Bearbeiten Informatiklehrkräfte außerhalb der Dienstzeit ein Blended-Learning-Modul für eine Lehrerfortbildung? - Flexibilität durch Online-Lernen vs. „Freizeit ist heilig“ .....</i>	367
<b>Marius Wehner, Alina Köchling</b> <i>Learning Analytics und Diskriminierung .....</i>	369

<b>Leonard Meyer, Thies Pfeiffer</b> <i>Virtual Reality-based Digital Reusable Learning Objects in Healthcare Training (ViRDiPA): 360°-Video Trainings</i> .....	371
<b>Annett Thüring, Kathrin Jäger</b> <i>Your Lesson, Your Way, Your Success</i> .....	373
<b>Demos</b>	
<b>Dana-Kristin Mah, Cornelia Gamst, Lavinia Ionica, Christian Dufentester</b> <i>Entwicklung des „KI-Campus – Die Lernplattform für Künstliche Intelligenz“</i> ....	377
<b>Ahmed Hassan, Niels Pinkwart</b> <i>'Zirkus Empathico 2.0' A serious mobile game for empathy enhancement in children with Autism</i> .....	379
<b>Sebastian Wollny, Jan Schneider, Marc Rittberger, Hendrik Drachsler</b> <i>Tabbi - A Tangible Interface for Educational Chatbots</i> .....	381
<b>Sarah Voß-Nakkour, Patrick Sacher, David Weiß, Thorsten Gattinger</b> <i>LernBar 4.6: Barrierearme, für Learning Analytics optimierte Web Based Trainings</i> .....	383
<b>Svenja Noichl, Ulrik Schroeder</b> <i>InfoBiTS - Informatische Bildung für Technikferne Seniorinnen und Senioren</i> .....	385
<b>Lubna Ali, Patrick Aufdermauer, René Röpke, Ulrik Schroeder</b> <i>Ein Webbasiertes Tool zur Konvertierung von Bildungsmaterialien in OER</i> .....	387
<b>Fiona Draxler, Elena Wallwitz, Albrecht Schmidt, Lewis L. Chuang</b> <i>An Environment-Triggered Augmented-Reality Application for Learning Case Grammar</i> .....	389
 <i>Verzeichnis der Autorinnen und Autoren</i> .....	 391



## **Keynotes**



## Current and Emerging Educational Realities - Shaping the Digital Transformation?

Sabine Seufert  <sup>1</sup>

**Abstract:** This paper gives an overview of the keynote and is structured in three parts: 1) What are current educational realities due to the COVID crises? 2) What are emerging trends and upcoming educational realities due to advanced technologies such as Artificial Intelligence (AI) and robotics? 3) How to shape digital transformation in educational institutions in order to exploit the potential of advanced technologies for desirable educational realities?

**Keywords:** Augmentation, Augmentation Strategies, Personalized AI-enabled learning, Learning Analytics, Immersive Learning, Conversational Learning


### 1 What are *current educational realities* due to the COVID crises?

In contrast to experiences that are planned from the beginning and designed to be online, emergency remote teaching is a temporary shift of instructional delivery to an alternate delivery mode due to crisis circumstances. The primary objective in these circumstances is not to re-create a robust educational ecosystem but rather to provide temporary access to instruction and instructional supports in a manner that is quick to set up and is reliably available during an emergency or crisis. The drastic shift towards emergency remote teaching might lead to the risk that the technology domain (focusing on operational information and communication technologies [ICT] skills) has gained strong momentum in schools as teachers receive tons of guidelines and technical advice. Students and teachers might be in different educational realities

### 2 What are *emerging trends and upcoming educational realities*?

To some extent, most of the students and teachers are already using advanced technologies, such as AI in their daily lives. According to Wahlster [Wa17] we distinguish first and second wave of digitalization. In the first wave of digitization, the focus is on ensuring that all relevant data is machine-readable and can therefore be digitally processed. The second wave of digitization is characterized by the understanding of digitally available data. It therefore outstrips the results of the first world, but unfolds a

---

<sup>1</sup> University of St.Gallen, Institute of Business Education and Educational Management, St.Jakob-Strasse 21, St.Gallen, CH-9000, sabine.seufert@unisg.ch,  <https://orcid.org/0000-0002-3807-6460>

much greater power of innovation, as it enables disruptive business models in a platform economy of self-learning smart services. Further advanced technologies such as robotics and immersive technologies could have further potential for educational purposes. Three trends seem to be emerging for educational realities:

- a) *Personalization*: AI-empowered adaptive learning support systems, personalized learning based on learning analytics in digital ecosystems [Se19];
- b) *Immersion*: Learning with immersive technologies such as Augmented Reality (AR) and Virtual Reality (VR) and learning in mixed realities,
- c) *Conversational Learning*: chatbots and social robots as new forms of human-computer interaction

### **3 How to shape *digital transformation* in educational institutions in order to exploit the potential of advanced technologies?**

There is a certain risk that the next wave will overrun schools even more than the first wave of digitalization. The danger is that investments made today in certain digital skills will be outdated in five years of time. New human-machine interactions raise new fundamental questions that cannot yet be answered comprehensively by teacher education and professional development of teachers. Therefore, it might be useful to better understand and further investigate the concept of augmented work and augmentation strategies for the teaching profession in order to identify teachers' technology-relevant knowledge, skills and attitudes [SJS20]. Drawing on Davenport and Kirby [DK16], five augmentation strategies may be particularly relevant for the teaching profession in order to manage AI transformation in educational institutions [SJS20].

## **Bibliography**

- [DK16] Davenport, T. H.; Kirby, J.: Just how smart are smart machines? In MIT Sloan Management Review, 2016, 57; pp. 21–25.
- [Se19] Seufert, S. et al.: A Pedagogical Perspective on Big Data and Learning Analytics: A Conceptual Model for Digital Learning Support. In Technology, Knowledge and Learning, 2019, 24; pp. 599–619.
- [SJS20] Sabine Seufert; Josef Guggemos; Stefan Sonderegger: Digitale Transformation der Hochschullehre: Augmentationsstrategien für den Einsatz von Data Analytics und Künstlicher Intelligenz, 2020.
- [Wa17] Wahlster, W.: Künstliche Intelligenz versus menschliche Intelligenz: Wie lernen, verstehen und denken Computer?, Mainz, 2017.



## Technology-Based Assessment in 21st Century Education

Samuel Greiff <sup>1</sup>

**Abstract:** This extended summary of a keynote given at the DELFI 2020 conference highlights current innovations in the field of educational assessment. It discusses three areas of rapid development: (1) the switch to technology in educational large-scale assessments; (2) the integration of assessment and learning in formal and informal learning contexts; (3) the role that new methods such as educational data mining and artificial intelligence play in educational assessment.


**Keywords:** Technology-Based Assessment; Large-Scale Assessment; Formative Assessment; Learning Analytics; Educational Data Mining; Artificial Intelligence

### 1 Educational Assessment: Recent Developments

Over the last years, educational assessments have undergone substantial changes. Several decades ago, educational tests were almost exclusively used in a summative way to evaluate the state of knowledge or proficiency that learners had acquired. With the advent of technology and with a broader conceptual understanding of assessment, the landscape has changed dramatically [IG19]. Whereas summative assessments remain an important part of testing, many assessment approaches are moving away from simple paper-pencil based assessments to integrated learning-assessment environments. This has had strong implications for students, teachers, institutions, and entire educational systems. On the basis of simulated assessment environments, news about student rankings, under- and overperforming countries, and novel ideas on how to improve educational systems are prominently featured in the media. While the scientific quality of these headlines vary considerably, there are interesting scientific, political, and practical developments and challenges happening in the field of educational assessment.

These developments are mostly driven by two concurrent pathways: (1) The development of quickly emerging technologies that allow for new measures related to the assessment of various skills such as analyses of process data and (2) through international large-scale assessments such as the Programme for International Student Assessment (PISA) commenced by the OECD [O09] that allow for empirically motivated comparisons of educational systems in a globalized world.

---

<sup>1</sup> University of Luxembourg, Department of Behavioural and Cognitive Sciences, 4366 Esch-sur-Alzette,  
samuel.greiff@uni.lu,  <https://orcid.org/0000-0003-2900-37340000>

## 2 Areas of Current Innovation

Three areas in educational assessment are particularly vibrant and likely to be associated with innovation in the near future: (1) the comprehensive switch to technology in educational large-scale assessments; (2) the integration of assessment and learning in formal and informal learning contexts; (3) the role that new methods such as educational data mining and artificial intelligence play in educational assessment.

With regard to (1), PISA that tests 15-year old students in over 70 countries worldwide was among the first large-scale assessments that began administering assessments through the computer in a large-scale setting. This has led to the availability of log-file data that introduces a new layer of understanding beyond the mere question whether an item was answered correctly or incorrectly by providing the chance to look at the actual behaviors underlying test-performance both on an individual student level but also on the level of entire educational systems [GWA15]. With regard to (2), assessment is nowadays increasingly understood as a learning experience that should provide an added value to the student who is being tested as well. Formative assessment is sometimes used to contrast summative assessment or, in a similar manner, assessment for learning is the new term that is contrasted with assessment of learning [Sh16]. With regard to (3) and closely associated with the previous point, new methods and even entirely new research areas such as the field of educational data mining are emerging that make use of algorithms in education. These algorithms often stem from computer science and are used to discover fuzzy relationships in large educational data sets to increase the understanding of learning.

As the field of educational assessment moves ahead along with the digital revolution, there will be many implications for educational policy and educational practice. Interdisciplinary research has a unique opportunity to contribute to these developments.

### Bibliography

- [GWA15] Greiff, S.; Wüstenberg, S.; Avvisati, F.: Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, 91, 92-105, 2015.
- [IG19] Iliescu, D.; Greiff, S.: The impact of technology on psychological testing in practice and policy. *European Journal of Psychological Assessment*, 35, 151-155, 2019.
- [O09] OECD: Programme for international student assessment: An overview. Paris: OECD, 2009.
- [Sh16] Shute, V. J. et al.: Advances in the science of assessment. *Educational Assessment*, 21, 1-27, 2016.

## **Forschungs- und Praxisbeiträge**



# Evaluation einer Augmented-Reality-Anleitung für eine komplexe Montageaufgabe

## Vergleich einer Smartphone-basierten Augmented-Reality-Anleitung mit einer klassischen Papieranleitung für die Anlernphase in der manuellen Montage

Johannes Funk<sup>1</sup> und Ludger Schmidt<sup>2</sup>

**Abstract:** Die vorliegende Studie vergleicht den Einsatz einer markerbasierten AR-Anleitung mit einer in der manuellen Montage üblichen Papieranleitung. Es wurden die Hypothesen geprüft, ob sich die Anleitungsart auf die Montagezeit, die Anzahl der Fehler, die Gebrauchstauglichkeit und die Beanspruchung der Mitarbeiter auswirkt. Statt mit studentischen Stichproben und künstlichen Montageaufgaben (z. B. Legobaugruppen) wurde die Studie mit 16 Auszubildenden an einem realen Arbeitsplatz für die Montage von Notgriffen zur Türentriegelung in Schienenfahrzeugen durchgeführt. Es wurden fünf Montagedurchgänge durchgeführt. Die Montagezeit und die gemachten Fehler wurden aus Videoaufzeichnungen ermittelt. Die Gebrauchstauglichkeit (SUS) und die Beanspruchung (NASA-TLX) wurden mit Fragebögen erhoben. Nachdem die AR-Gruppe zu Beginn langsamer montierte als die Papier-Gruppe, war diese im fünften Durchgang signifikant schneller. Die vergleichbare Fehleranzahl, Gebrauchstauglichkeit und Beanspruchung machen markerbasierte AR-Anwendungen besonders durch den leichten Einstieg und die geringen Kosten interessant für die Wissensvermittlung in der manuellen Montage.

**Keywords:** Augmented Reality, Digitalisierung, Montage, Anlernen, Feldstudie

## 1 Einleitung

Neben erheblicher Chancen für die Wirtschaft und den Mittelstand durch die zunehmende Digitalisierung [De16, In19] ergeben sich auch grundlegende Veränderungen der Arbeitswelt [Ap18]. Die voranschreitende Globalisierung, Änderungen der Markanforderungen und individuelle Kundenbedürfnisse sorgen für eine Zunahme der Variantenvielfalt und eine immer komplexer werdende Produktionsorganisation [WKG04]. Eine wachsende Produktindividualität führt zu sinkenden Losgrößen und kürzeren Produktlebenszyklen [Sc15]. Die dadurch häufig wechselnden Aufgabenfelder führen bspw. zu einer erhöhten Häufigkeit, mit der sich Mitarbeiter mit neuen Inhalten auseinandersetzen müssen [Hö17]. Aufgrund der neuen Herausforderungen gewinnt die Flexibilität, die nötig ist, um sich auf neue Varianten einzustellen, immer stärker an Bedeutung [Be98]. Hierdurch wird häufig auf eine manuelle oder teilautomatisierte

---

<sup>1</sup> Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstraße 7, 34125 Kassel, J.Funk@uni-kassel.de

<sup>2</sup> Universität Kassel, Fachgebiet Mensch-Maschine-Systemtechnik, Mönchebergstraße 7, 34125 Kassel, L.Schmidt@uni-kassel.de

Montage gesetzt, um die hohe Flexibilität des Menschen zu nutzen [Sc15]. Die hohe Variantenvielfalt hat in der Montage eine Zunahme von Bauteilen und Werkzeugen zur Folge, was eine hohe Anforderung der Mitarbeiter bedeutet [Fr98]. Für die Montage von Baugruppen stehen Mitarbeitern oft nur Montagepläne in Form von Zusammenbauzeichnungen oder Arbeitsanweisungen in Textform zur Verfügung, deren Verständlichkeit in vielen Fällen unzureichend ist [Sc15]. Hinzu kommt, dass immer öfter ungelernete Mitarbeiter und Mitarbeiter mit geringen Sprachkenntnissen in Betrieben arbeiten, wodurch Fehler durch bspw. falsches Lesen oder falsches Interpretieren von Arbeitsunterlagen entstehen [Rö15].

Ein Ansatz zur Verbesserung der Verständlichkeit von Arbeitsanweisungen und Montageanleitungen können innovative Arten der Wissensvermittlung mit Augmented Reality (AR) und Virtual Reality (VR) darstellen [Sc04]. Der Begriff Augmented Reality beschreibt die Ergänzung oder auch Erweiterung der Realität um virtuelle Inhalte [Dö13], die in Echtzeit, interaktiv und räumlich korrekt registriert werden [Az97]. Neben der Darstellung von virtuellen Inhalten bspw. auf einem transparenten Bildschirm eines Head-Mounted Display (Optical See-Through) [Tö10], können Bildschirme verwendet werden, um ein Kameravideo mit Inhalten zu erweitern (Video See-Through) [Cu16]. In Kombination mit dem optischen Trackingverfahren des Markertrackings lassen sich einfache AR-Anwendungen erstellen, die bspw. mit dem Smartphone oder einem Tablet verwendet werden können. Neben schwarz-weißen Mustern können auch beliebige andere Bilder als Marker verwendet werden, die von der Kamera des verwendeten Geräts erkannt und verwendet werden, um virtuelle Inhalte in Bezug zum jeweiligen Marker lagerichtig einzublenden [Tö10]. Da neben einer Kamera zum Erfassen der Marker nur meist kostenfreie Software zum Programmieren benötigt wird, stellt markerbasiertes AR eine kostengünstige Option für den Einsatz von AR dar [Tö10]. Hinzu kommt der einfache Umgang mit Markern, der auch für Laien einfach zu erlernen ist. Die Marker können schnell erstellt, ausgedruckt und anschließend frei platziert werden [Dö13].

Der Einsatz von AR-Markern in Montageanleitungen wurde bereits in verschiedenen Studien untersucht. Hou und Wang fanden heraus, dass Probanden mit einer AR-Montageanleitung eine Legobaugruppe schneller und mit weniger Fehlern montierten als mit einer gedruckten Papieranleitung [HW13]. Loch et al. untersuchten ebenfalls die Montage einer Legobaugruppe, verglichen hierbei aber eine Videoanleitung mit einer AR-Anleitung. Die Studie mit 17 Probanden zeigte keinen signifikanten Unterschied in der Montagedauer und der Beanspruchung (NASA-TLX), allerdings wurden in der AR-Gruppe signifikant weniger Fehler gemacht [LQB16]. Da in der Studie beide Gruppen verschiedene Legomodelle aufbauten, ist eine Vergleichbarkeit der Komplexität der Montageaufgabe fraglich.

Statt einer Legobaugruppe betrachtete Hořejší die Montage eines Siphons und verglich hierbei ebenfalls eine Montageanleitung in Papierform mit einer AR-Anleitung. Die AR-Anleitung zeigte dem Probanden neben einem 3D-Modell des aktuellen Bauteils ebenfalls das korrekte Teilelager mit Pfeilen an, in dem das benötigte Bauteil zu finden war. Die

Studie zeigte eine große Zeitersparnis in der Montage des ersten Montagedurchgangs [Ho15]. Dieser könnte allerdings mit der fehlenden Unterstützung in der Teileauswahl aus den Teilelagern der Papieranleitung zusammenhängen.

Die vorgestellten Studien betrachten Montageaufgaben, die teilweise stark von tatsächlicher Montage in Unternehmen abweichen. Um eine Übertragbarkeit der Laborstudien auf eine reale Montageaufgabe in Unternehmen zu gewährleisten, wird nachfolgend die vergleichende Evaluation einer Papieranleitung und einer AR-Anleitung an einem realen Montagearbeitsplatz für die Montage von Notgriffen zur Türentriegelung in Schienenfahrzeugen untersucht. Hierbei liegt der Fokus auf dem Einfluss der Anleitungsart auf die Montagedauer, die dabei gemachten Fehler, die Gebrauchstauglichkeit und die Beanspruchung der Mitarbeiter. In der Studie werden folgende Hypothesen geprüft: Die Anleitungsart wirkt sich auf die Montagezeit (H1) und die Anzahl der Montagefehler (H2) aus. Die AR-Anleitung unterscheidet sich in der Gebrauchstauglichkeit von der klassischen Papieranleitung (H3) und die Anleitungsart wirkt sich auf die Beanspruchung der Mitarbeiter aus (H4).

Die Besonderheit der Studie liegt in dem realen Anwendungsfall, da diese am Montageplatz eines Industrieunternehmens mit einer realer Baugruppe durchgeführt wird. Im Gegensatz zu häufig eingesetzten studentischen Stichproben wird die Studie mit Auszubildenden des Unternehmens durchgeführt, die sich alle in einer technischen Ausbildung befinden und vertraut mit der manuellen Montage sind. Hierdurch entsprechen die Probanden der Zielgruppe, für die der mögliche Einsatz von AR-Anleitungen in der Montage interessant ist.

## **2 Methode**

### **2.1 Montageaufgabe**

In der Studie wird die manuelle Montage eines Notgriffs zum Öffnen von Bahntüren betrachtet. Die Baugruppe wird in bis zu 30 verschiedenen Varianten gefertigt. Der in der Studie verwendete Notgriff besteht aus 16 Bauteilen, und die Montage unterteilt sich in verschiedene Tätigkeiten, wie dem Positionieren, Verschrauben, Verpressen, Einhängen und Ausrichten von Bauteilen. In Abb. 1 ist links ein Modell des fertig montierten Notgriffs zu sehen.

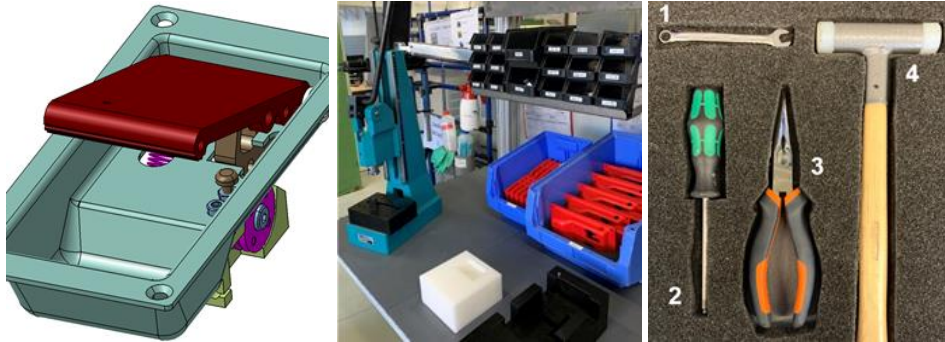


Abb. 1: 3-D-Modell des zu montierenden Notgriffs (links); Montagearbeitsplatz mit Teilelagern, Montagevorrichtungen und eine Einpressvorrichtung (mittig); für die Montage benötigtes Werkzeug (rechts): Ring-Maulschlüssel SW 8 (1), Innensechskantschlüssel SW 4 (2), Spitzzange (3), Kunststoffhammer (4)

Die Montage findet stückweise an einem Einzelarbeitsplatz statt. Trotz der komplexen Aufgabe kann die Montage von ungelerten Personen nach kurzer Anlernzeit selbstständig durchgeführt werden. Der Arbeitsplatz mit einem Montagetisch (90 cm x 150 cm) verfügt über Teilelager, die mit Positionsnummern gekennzeichnet sind. Als Montagehilfe stehen Montagevorrichtungen bereit, in die Bauteile eingelegt werden können und so die korrekte Positionierung erleichtern. Neben einfachen Handwerkzeugen steht eine Einpressvorrichtung für das Einpressen von Metallstiften zur Verfügung. Der Arbeitsplatz und die benötigten Werkzeuge sind in Abb. 1 zu sehen.

## 2.2 AR-Montageanleitung

Für die Studie wurde eine markerbasierte AR-Montageanleitung für den Notgriff als Smartphone-Anwendung erstellt. Als Grundlage diente hierzu eine bestehende und im Unternehmen eingesetzte Papieranleitung. Die Inhalte der Papieranleitung wurden in die AR-Anwendung übertragen. Für die AR-Anwendung wurde der Montageprozess in sieben Teilschritte unterteilt und mit je einem Marker versehen. Diese Marker werden von der Anwendung erkannt und verwendet, um passende Informationen lagerichtig anzuzeigen. In jedem der sieben Teilschritte werden dem Mitarbeiter 3-D-Modelle der gerade benötigten Bauteile, ein Textfeld mit den dazugehörigen Positionsnummern der Teilelager sowie ggf. weitere Informationen über den auszuführenden Arbeitsschritt angezeigt. Teilweise bestehen die Teilschritte aus mehreren aufeinander aufbauenden Montageschritten. Hierzu wurde die Anwendung um zwei Schaltflächen zum Vor- und Zurückschalten der angezeigten Informationen erweitert. Die Nutzung der Anwendung sowie die Darstellung auf dem Smartphone ist in Abb. 2 zu sehen.



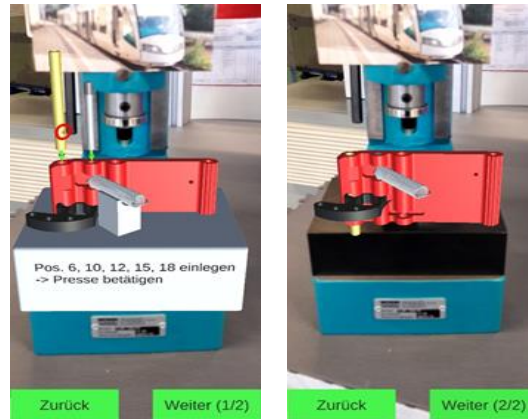


Abb. 2: Smartphone-Ansicht für Teil eins von zwei mit Angabe der Teilpositionsnummern sowie der Arbeitsaufforderung „Presse betätigen“, Schaltflächen „Zurück“ und „Weiter“ zum selbständigen Durchblättern von Informationen (links); Anzeige der korrekt montierten Teile (rechts)

Die AR-Anwendung wurde in der 3-D-Entwicklungsumgebung Unity (Version 2019.1.12f1) und dem Vuforia Kit für markerbasiertes AR erstellt. In Unity wurden die 3D-Modelle der Baugruppe mit den Markern der entsprechenden Teilschritte verknüpft und teilweise mit Text für die Positionsnummer und weiteren Hinweisen ergänzt.

### 2.3 Messung

Die Technikaffinität wurde mit dem standardisierten TA-EG-Fragebogen mit 19 Items auf einer fünfgliedrigen Likert-Skala von 1 („Trifft gar nicht zu“) bis 5 („Trifft voll zu“) erhoben [Ka09]. Nachdem die Werte negativ formulierter Items negiert wurden, werden die Mittelwerte der vier Skalen *Begeisterung*, *Kompetenz*, *negative und positive Einstellung* zu einem Gesamtwert gemittelt. Die Werte liegen in einem Bereich von 1 (sehr niedrige Technikaffinität) und 5 (sehr hohe Technikaffinität).

Die Durchführzeiten und die Anzahl der Montagefehler wurden mittels Videoaufzeichnung für jeden der fünf Montagedurchgänge ausgewertet.

Zur Messung der Gebrauchstauglichkeit wurde der System-Usability-Scale-Fragebogen (SUS) verwendet [Br14]. Dieser besteht aus zehn Items, von denen jeweils fünf positiv und negativ formuliert sind. Die Items sind auf einer fünfgliedrigen Likert-Skala von 1 („Ich stimme gar nicht zu“) bis 5 („Ich stimme voll zu“) auszufüllen. Entsprechend den Vorgaben von Brooke werden für den Gesamtwert die negativ formulierten Items negiert, anschließend werden alle Werte von 1-5 auf 0-4 umgerechnet, summiert und mit dem Faktor 2,5 multipliziert [Br14]. Hierdurch ergibt sich ein möglicher Gesamtwert von 0 (sehr schlechte Gebrauchstauglichkeit) und 100 (sehr gute Gebrauchstauglichkeit).

Die Beanspruchung wurde mit dem NASA-TLX-Fragebogen erhoben [HS88]. Dieser besteht aus der subjektiven Bewertung der sechs Unterpunkte der geistigen, körperlichen und zeitlichen Anforderung sowie Leistung, Anstrengung und Frustration auf einer 20-gliedrigen Skala von „gering“ bis „hoch“. In Paarvergleichen wählt der Teilnehmer aus je zwei der sechs Unterpunkte das aus, was für ihn bei der Arbeitsaufgabe bedeutsamer war. Hierdurch kann der Gesamtwert gewichtet werden. Der Gesamtwert liegt zwischen 0 (keine Beanspruchung) und 100 (extreme Beanspruchung).

#### **2.4 Versuchsaufbau**

Für die AR-Anwendung wurde ein Smartphone Galaxy S8 mit einer Displaygröße von 5,8" (14,7 cm) und einer Auflösung von 1440 x 2960 Pixeln verwendet. Die Kamera mit 12 Megapixeln erkennt die am Montageplatz befestigten Marker und blendet lagerichtige Inhalte in das auf dem Display angezeigte Kamerabild. Für die Vergleichsgruppe mit Papieranleitung wurde keine zusätzliche Technik benötigt.

#### **2.5 Versuchsablauf**

Zu Beginn wurden die Probanden durch das Ziehen eines von 16 vorbereiteten Losen (acht für jede der beiden Gruppen) entweder der *Papieranleitung* (Papier-Gruppe) oder *AR-Anwendung* (AR-Gruppe) zugeteilt. Anschließend erhielten die Probanden eine Versuchseinführung mit Informationen über die Studie und deren Ablauf sowie eine Einverständniserklärung und einen demografischen Fragebogen (Alter, Geschlecht, Ausbildungsberuf). Die Einführungsunterlagen gaben keinen Informationen über die Anleitungsart der jeweils anderen Gruppe. Die Gruppe *AR-Anwendung* wurde zusätzlich nach Vorerfahrungen mit AR-Anwendungen gefragt. In dieser Gruppe endete die Vorbereitung mit Informationen zu Vuforia-Markern und dem praktischen Ausprobieren einer Beispiel-AR-Anwendung. Nach der Klärung offener Fragen begann die Montageaufgabe.

Die Aufgabe bestand aus der Montage von fünf Notgriffen. Abb. 3 zeigt Probanden bei der Montage mit der Papieranleitung (links) und der AR-Anleitung (rechts).



Abb. 3: Probanden während der Montage des Notgriffs mit Papieranleitung (links) und AR-Anleitung (rechts)

Nach jedem Durchgang gab es eine Pausenzeit von etwa zwei Minuten, in denen der Versuchsleiter den Montagearbeitsplatz und die AR-Anwendung in den Ausgangszustand brachte. Während der Montage wurden dem Probanden gemachte Fehler durch den Versuchsleiter mithilfe von zwei entsprechenden Karten „falsches Bauteil verwendet“ (Fehler 1) und „Bauteil falsch montiert“ (Fehler 2) angezeigt.

Nach Abschluss des fünften Durchgangs füllten die Probanden die Fragebögen (TA-EG, SUS, NASA-TLX) aus und hatten die Möglichkeit, Feedback zur verwendeten Anleitung zu geben.

## 2.6 Stichprobe

Zur Überprüfung der Hypothesen wurde eine Feldstudie in einem Industrieunternehmen für die Herstellung von elektrischen und mit Druckluft betriebenen Türsystemen durchgeführt. An der Studie nahmen 16 männliche Probanden des Unternehmens mit einem Durchschnittsalter von 19,4 Jahren ( $SD = 2,2$  Jahre) teil. Ein zweiseitiger t-Test zeigt sowohl beim Alter als auch der Technikaffinität keine signifikanten Unterschiede der Gruppen *Papieranleitung* (Alter:  $M = 19,3$  Jahre;  $SD = 1,7$  Jahre; TA-EG:  $M = 3,84$ ,  $SD = 0,29$ ) und *AR-Anwendung* (Alter:  $M = 19,6$  Jahre;  $SD = 2,7$  Jahre; TA-EG:  $M = 3,86$ ,  $SD = 0,45$ ) (Alter:  $t(14) = 0,33$ ,  $p = 0,745$ ; TA-EG:  $t(14) = 0,10$ ,  $p = 0,918$ ). Die Teilnehmer befanden sich alle in der Ausbildung zum Industriemechaniker und besaßen daher alle Vorerfahrung in der manuellen Montage von komplexen Baugruppen.

## 3 Ergebnisse

Zu Beginn der Auswertung wurden alle Skalen mit dem Shapiro-Wilk-Test auf Normalverteilung und mit dem Levene-Test auf Varianzhomogenität geprüft. Bis auf die

Anzahl gemachter Fehler weisen alle Werte auf den nachfolgend vorgestellten Skalen eine Normalverteilung und Varianzhomogenität auf.

Bei Hypothese H1 wird geprüft, ob sich die Durchführzeit der Montage zwischen den Gruppen Papier und AR unterscheidet. Die aus den Videos ermittelten Montagezeiten sind in Abb. 4 zu sehen.

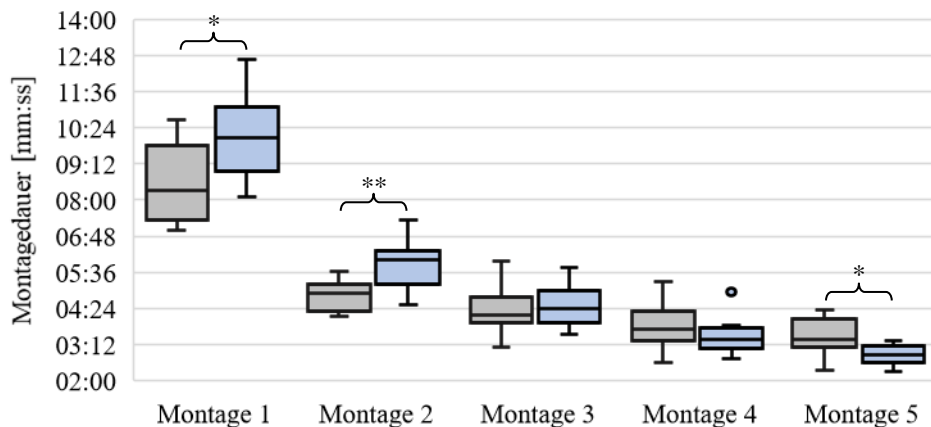


Abb. 4: Boxplots der Montagezeiten der Papier- (grau) und AR-Gruppe (blau) gruppiert nach den Montagedurchgängen

Zur Prüfung von H1 wurde für jeden Montagedurchgang ein zweiseitiger t-Test für unabhängige Stichproben berechnet. In Tab. 1 sind die Mittelwerte, Standardabweichungen und die Ergebnisse der t-Tests angegeben.

Montagedurchgang	Papier-Gruppe	AR-Gruppe	t-Test
Montage 1	M = 8,53 min, SD = 1,31 min	M = 10,10 min, SD = 1,44 min	$t(14) = 2,28,$ $p = 0,039$
Montage 2	M = 4,82 min, SD = 0,52 min	M = 5,89 min, SD = 0,85 min	$t(14) = 3,04,$ $p = 0,009$
Montage 3	M = 4,36 min, SD = 0,84 min	M = 4,46 min, SD = 0,70 min	$t(14) = 0,27,$ $p = 0,789$
Montage 4	M = 3,80 min, SD = 0,81 min	M = 3,50 min, SD = 0,66 min	$t(14) = -0,81,$ $p = 0,431$
Montage 5	M = 3,46 min, SD = 0,66 min	M = 2,86 min, SD = 0,34 min	$t(14) = -2,32,$ $p = 0,036$
Gesamtdauer	M = 24,98 min, SD = 3,43 min	M = 26,81 min, SD = 3,00 min	$t(14) = 1,14,$ $p = 0,274$

Tab. 1: Mittelwerte und Standardabweichungen der Montagezeiten der Papier- und AR-Gruppe über fünf Durchgänge sowie die Ergebnisse der t-Tests

Bei Hypothese H2 wird die Anzahl der Fehler während der Montage bei den beiden Anleitungsarten geprüft. Hierzu wurden die Videoaufzeichnungen ausgewertet und die gemachten Fehler pro Proband und Montagedurchgang betrachtet. Die Papier-Gruppe machte insgesamt im Durchschnitt 3,25 Fehler ( $SD = 2,49$  Fehler) und die AR-Gruppe 3,12 Fehler ( $SD = 1,13$  Fehler). Abb. 5 zeigt die durchschnittliche Anzahl gemachter Fehler pro Montagedurchgang nach Gruppen. Mann-Whitney-U-Tests zeigten keine signifikanten Unterschiede.

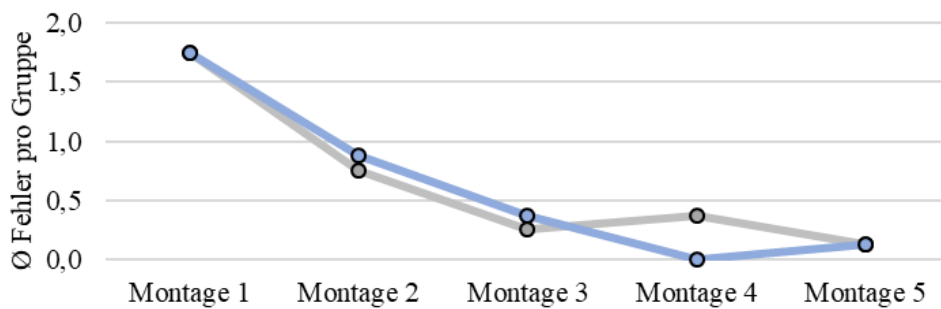


Abb. 5: Durchschnittliche Montagefehler der Papier- (grau) und AR-Gruppe (blau) über den Montagedurchgängen

Bei den Hypothesen H3 und H4 wird geprüft, ob sich die Gebrauchstauglichkeit (H3) und die Beanspruchung (H4) bei der Nutzung der AR-Anwendung von der der Papieranleitung unterscheidet. Die beiden Ergebnisse sind in Abb. 6 angegeben.

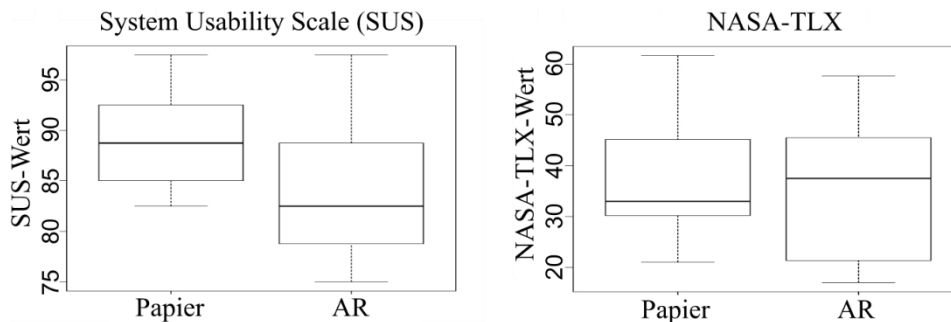


Abb. 6: Gebrauchstauglichkeit (SUS) (links) und der Beanspruchung (NASA-TLX) (rechts)

Zur Prüfung auf Unterschiede wurde jeweils ein zweiseitiger t-Test für unabhängige Stichproben berechnet. Die mit dem SUS durchschnittlich erhobene Gebrauchstauglichkeit ist in der Papier-Gruppe ( $M = 89,06$ ,  $SD = 5,16$ ) höher als in der AR-Gruppe ( $M = 84,06$ ,  $SD = 7,31$ ). Dieser Unterschied ist jedoch nicht signifikant ( $t(14) = -1,58$ ,  $p = 0,137$ ).

Die mit dem NASA-TLX erhobene durchschnittliche Beanspruchung liegt bei der AR-Gruppe ( $M = 35,41$ ,  $SD = 14,71$ ) niedriger als in der Papier-Gruppe ( $M = 37,42$ ,

$SD = 13,59$ ). Dieser Unterschied ist jedoch ebenfalls nicht signifikant ( $t(14) = -0,28$ ,  $p = 0,780$ ).

## 4 Diskussion und Fazit

Die Studie zeigt gemischte Ergebnisse der Montagezeiten. In den Durchgängen eins und zwei sind die Probanden der Papier-Gruppe signifikant schneller als die der AR-Gruppe. Dieses Ergebnis weicht von den Befunden von Hořejší ab, wo besonders im ersten Montagedurchgang eine hohe Zeitersparnis mit der AR-Anleitung erreicht werden konnte [Ho15]. Dieser Unterschied könnte daran liegen, dass in der vorliegenden Studie versucht wurde, die Inhalte beider Anleitungen identisch zu halten, wohingegen Hořejší der AR-Gruppe zusätzliche Unterstützung in der Teileauswahl aus den Teilelagern gab. Nachdem keine Unterschiede der Zeiten für die Durchgänge drei und vier vorliegen, ist die AR-Gruppe in Durchgang fünf signifikant schneller als die Papier-Gruppe. Es ist zu vermuten, dass die anfänglich guten Zeiten der Papier-Gruppe trotz des Übens mit einer Beispiel-AR-Anwendung durch die Vorerfahrung der Probanden im Umgang mit der Papieranleitung stammen. Die AR-Gruppe musste sich hingegen doch noch mit der neuen Anleitungsart und mit deren Umgang während der Montage vertraut machen. Der Verlauf der Zeiten und die schnellere Montage in Durchgang fünf deuten dennoch auf das allgemeine Potential von AR-Anleitungen.

Die Ergebnisse lassen bezüglich Fehlern keine eindeutige Aussage zu. Die beiden Gruppen unterscheiden sich weder in einzelnen Durchgängen, noch in der Betrachtung der Gesamtfehleranzahl. Beide Gruppen machen sehr wenige Fehler, wobei die gemachten Fehler häufig auf Unachtsamkeit wie dem Vergessen der Montage kleiner Bauteile oder dem Prüfen von Feinheiten (z. B. korrekte Ausrichtung der Wellen) zurück zu führen sind.

Die Gebrauchstauglichkeit beider Anleitungen liegt mit einem Durchschnittswert von 89,06 (Papier-Gruppe) bzw. 84,06 (AR-Gruppe) knapp über bzw. unter einer Bewertung von *exzellent* (bei 85) [BKM09]. Damit liegen beide in einem sehr guten Bereich, Unterschiede in der Gebrauchstauglichkeit konnten nicht nachgewiesen werden. Die subjektive Beanspruchung der Mitarbeiter durch die Anleitungsart liegt mit durchschnittlich 37,42 (Papier-Gruppe) und 35,41 (AR-Gruppe) ebenfalls auf einem ähnlichen Niveau und weist keine signifikanten Unterschiede auf. Um die Werte einzuordnen, wird auf Ergebnisse anderer Studien zurückgegriffen. In einer Meta-Analyse wurden die NASA-TLX-Werte von 237 wissenschaftlichen Studien mit insgesamt 1173 Datensätzen ausgewertet. Der Median aller Werte liegt bei 49,93, 25 % der Werte liegen unter 36,77 [Gr15]. Damit lassen sich die vorliegenden Werte beide einer eher geringen Beanspruchung zuordnen.

Obwohl in den Versuchsunterlagen bewusst auf Informationen über die Anleitungsart der jeweils anderen Versuchsgruppe verzichtet wurde, lässt sich nicht ausschließen, dass

Probanden darüber Bescheid wussten. Mögliche Verzerrungseffekte durch eine höhere Motivation der AR-Gruppe sind daher möglich.

Die Ergebnisse zeigen teilweise Überschneidungen, aber auch Unterschiede zu ähnlichen Studien mit studentischen Stichproben und künstlichen Montageaufgaben. Hier wird die Bedeutung spezieller Stichproben deutlich, um die Übertragbarkeit von Ergebnissen zu ermöglichen. Die vergleichbare hohe Gebrauchstauglichkeit und die niedrige Beanspruchung machen markerbasierte AR-Anwendungen besonders durch den leichten Einstieg und die geringen Kosten interessant für die manuelle Montage. Zwar konnten in dieser Studie keine positiven Effekte auf die Montagezeiten und die Fehler nachgewiesen werden, allerdings kann weitere Forschung helfen, den Einfluss von AR auf die Montage genauer zu verstehen und gezielt einzusetzen.

### Literaturverzeichnis

- [Ap18] Apt, W. et al.: Einsatz von digitalen Assistenzsystemen im Betrieb, 2018.
- [Az97] Azuma, R. T.: A Survey of Augmented Reality. In *Presence: Teleoperators and Virtual Environments*, 6, S. 355–385, 1997.
- [Be98] Behr, P.: Komplexitätsmanagement von Produkten und Prozessen - Wie lässt sich Komplexität managen? In (VDI-Gesellschaft Entwicklung, Konstruktion, Vertrieb Hrsg.): *Effektive Entwicklung und Auftragsabwicklung variantenreicher Produkte*. VDI-Verlag, Düsseldorf, S. 39–51, 1998.
- [BKM09] Bangor, A.; Kortum, P.; Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. In *Journal of Usability Studies*, 4, S. 114–123, 2009.
- [Br14] Brooke, J.: SUS: a "quick and dirty" usability scale. In (Jordan, P. W. et al. Hrsg.): *Usability Evaluation in Industry*. Chapman and Hall/CRC, Boca Raton, S. 189–194, 2014.
- [Cu16] Curran, C.: Breakthroughs in optics that are reshaping augmented reality. <http://usblogs.pwc.com/emerging-technology/breakthroughs-in-optics-that-are-reshaping-augmented-reality/>, 16.10.2019.
- [De16] Demary, V. et al.: *Digitalisierung und Mittelstand. Eine Metastudie*. Institut der deutschen Wirtschaft Medien, Köln, 2016.
- [Dö13] Dörner, R. et al.: *Virtual und Augmented Reality (VR / AR)*. Springer, Berlin, Heidelberg, 2013.
- [Fr98] Franke, H.-J.: Produkt-Variantenvielfalt - Ursachen und Methoden zu ihrer Bewältigung. In (VDI-Gesellschaft Entwicklung, Konstruktion, Vertrieb Hrsg.): *Effektive Entwicklung und Auftragsabwicklung variantenreicher Produkte*. VDI-Verlag, Düsseldorf, S. 1–13, 1998.

- [Gr15] Grier, R. A.: How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59; S. 1727–1731, 2015.
- [Ho15] Hořejší, P.: Augmented Reality System for Virtual Training of Parts Assembly. In Procedia Engineering, 100; S. 699–706, 2015.
- [Hö17] Höhne, B. P. et al.: Agiles Lernen am Arbeitsplatz – Eine neue Lernkultur in Zeiten der Digitalisierung. In Zeitschrift für Arbeitswissenschaft, 71; S. 110–119, 2017.
- [HS88] Hart, S. G.; Staveland, L. E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research: Human Mental Workload, S. 139–183, 1988.
- [HW13] Hou, L.; Wang, X.: A study on the benefits of augmented reality in retaining working memory in assembly tasks. A focus on differences in gender. In Automation in Construction, 32, S. 38–45, 2013.
- [In19] D21 DIGITAL INDEX 2018/2019 Jährliches Lagebild zur Digitalen Gesellschaft. Hochschule der Medien, Stuttgart, 2019.
- [Ka09] Karrer, K. et al.: Technikaffinität erfassen – der Fragebogen TA-EG. In Der Mensch im Mittelpunkt technischer Systeme; S. 196–201, 2009.
- [LQB16] Loch, F.; Quint, F.; Brishtel, I.: Comparing Video and Augmented Reality Assistance in Manual Assembly. In (IEEE Hrsg.): 12th International Conference on Intelligent Environments. IEEE, London; S. 147–150, 2016.
- [Rö15] Röhrborn, S.: Deutsche Sprache, schwere Sprache? <https://www.vanguard.de/blog/deutsche-sprache-schwere-sprache>, 14.11.2019.
- [Sc15] Schenk, M.: Produktion und Logistik mit Zukunft. Springer, Berlin, Heidelberg, 2015.
- [Sc04] Schmidt, L. et al.: Augmented Reality als neue Form der Mensch-Technik-Interaktion. In (Luczak, H. et al. Hrsg.): Benutzerzentrierte Gestaltung von Augmented-Reality-Systemen. Düsseldorf: VDI-Verlag, Reihe 22, Mensch-Maschine-Systeme, S. 1–14, 2004.
- [Tö10] Tönnis, M.: Augmented Reality. Springer, Berlin, Heidelberg, 2010.
- [WKG04] Wiendahl, H.-P.; Gerst, D.; Keunecke, L.: Variantenbeherrschung in der Montage. Konzept und Praxis der flexiblen Produktionsendstufe. Springer, Berlin, Heidelberg, 2004.



## Potential vs. Practice: Challenges for the Implementation of Augmented Reality for Learning and Training in Practice

Niklas Osmers<sup>1</sup> and Michael Prilla<sup>2</sup>

**Abstract:** Augmented Reality (AR) holds many potentials for learning and training in organizations. However, there are surprisingly few examples that show an implementation of these potentials in practice. Motivated by this insight and similar observations through discussions with practitioners, we conducted an interview study to explore the usage of AR for learning and training in practice. Our results show that there is little systematic and established use in practice because of several challenges inhibiting its use. Our paper reports on these and discusses how they might be overcome.

**Keywords:** Augmented Reality, Learning, Training, Interview Study, Challenges, Practice

### 1 Introduction

Augmented Reality (AR) superimposes digital content such as objects, text or other items onto the real environment [Az97]. By using AR technology, users find themselves in a mixed reality [MK94] with an environment made up of both, real and virtual components. AR has been adopted in many different sectors of work like manufacturing [Ab17], service and maintenance [Fa16], healthcare [Si17], sports [Pa18] and many others. One specific area of interest is its potential support for learning and training in organizations. While its potential for learning and training *at work* has been recognized (e.g., [Di13, Fu17, Ho17, Sa16]), there are surprisingly few reports on implementations of these potentials in practice. Motivated by this, we asked ourselves whether and to which extent AR is currently implemented in practice as a learning and training tool as well as what success factors or challenges exist.

The work presented here aims at answering these questions. We conducted an interview study with 44 representatives of organizations interested in using AR for learning and training, service providers offering corresponding solutions and researchers engaged in projects with that purpose. We surveyed them about current and former projects as well as factors that led to success or provided challenges. Most examples we were told about were in rather immature states, and we identified many challenges for the integration of AR into learning and training processes. By reporting on the gap between the potential of AR for learning and training and current implementations, our work contributes to the

---

<sup>1</sup> Technical University of Clausthal, Human-Centered Information Systems, Julius-Albert-Str. 4, Clausthal-Zellerfeld, 38678, Niklas.Osmers@tu-clausthal.de

<sup>2</sup> Technical University of Clausthal, Human-Centered Information Systems, Julius-Albert-Str. 4, Clausthal-Zellerfeld, 38678, Michael.Prilla@tu-clausthal.de

knowledge on implementing AR and tapping into its potential for learning and training.

## **2 Literature Review: AR for Learning and Training at Work**

The potential of AR for learning is discussed by many researchers and practitioners. Compared to other media for learning, a substantial amount of this discussion has been held in the context of learning at work (as opposed to educational settings). There are many studies that show the applicability and the potential of AR for learning that is integrated into the workplace [Li17, Ni17, U116]. Among the available studies, AR has been shown to improve attentiveness, motivation and internalization of knowledge [Di13, Sa16], to ease the learning of standard procedures [Si17], to help people reproduce knowledge later on [Fu12], and to increase the confidence about one's own learning [Di13]. One specific promise that AR holds for learning is to bridge the gap between the (often theoretical) learning and the (practical) application of what has been learned [Ho17], which is crucial for the transfer of learning into practice [Ki75, Ph96]. As a consequence, AR is often used for learning or supporting practical work tasks [Ab17, Bü17, Mo17]. By using it for tasks like vocabulary training [Sa16] or simulating medical tasks [Az18, Si17], AR support more sustainable learning [Bo14].

Recent work on the application of AR for learning shows mixed results. Funk et al. [Fu17] found that projection-based AR is helpful for learning manual tasks. Radu and Schneider [RS19] show that learners arrive at a better understanding of physics when using AR. While these studies stress the potential of AR for learning, other work has arrived at less encouraging results. For example, Wilschut et al. [Wi19] found no significant differences in time or learning effects when comparing screen instructions to AR projection-based instructions for manual assembly tasks. Also, a study of Büttner et al. [Bü20] finds AR to be inferior to face-to-face trainings, only showing an advantage of preventing mislearning when compared to paper manuals. Werrlich et al. [We18] report similar findings.

Recent studies also point to the fact that despite its potential, there are only very few implementations of AR based learning and training in practice [Du18], and that its success and acceptance depends much on the (perceived) value added by such training [Ko18] as well as on the implementation of AR along established learning theories [Zh14].

## **3 Interview Study: Challenges of AR for Learning and Training in German-speaking countries**

Looking at the potential benefits of AR for learning and training portrayed in the previous chapter, one may assume that AR is already an established tool for learning and training in practice, or at least rapidly emerging for these purposes. Surprised by a lack of evidence and only very few insights into practical applications, we wanted to investigate the actual use of AR for training and learning in practice including benefits, potentials, disadvantages

and challenges of its implementation. In order to acquire an overview over the current situation and its state of implementation, we conducted an interview study with different actors that have some experience on applying AR for learning and training. In the interview, we asked questions based on the following research questions:

1. How is AR being used in practice for learning and training?
2. What challenges are being faced regarding this usage in practice?

We restricted our study to German-speaking companies in order to simplify differences.

### 3.1 Study Design

#### Procedure

Based on a thorough internet research, 255 German-speaking actors were identified to have some experience in using AR for learning and training. These contacts were investigated in order to determine their suitability and willingness to participate in our study. Out of these, we interviewed a total of 44 representatives from companies out of various fields of work (e.g. automotive or education). From July 2018 to December 2018, each interviewee was called for a 45 minutes long semi-structured interview by one of three researchers. Afterwards, all recorded audio files were transcribed and analyzed using a mixed content coding approach. First, three researchers marked all passages in line with our research questions with associated codes. Second, we reviewed these codes in detail with an inductive bottom-up method and analyzed resulting sub-codes.

#### Participants

Taking a closer look at the actors that were interviewed, we observed three major groups differing based on their perspective on AR as part of learning and training:

- **Service providers** use AR to support training and teaching services they offer to other companies. Though, they rarely use the technology for their own employees. (n = 17)
- **Users** apply AR tools for teaching and training their own employees. (n = 15)
- **Researchers** are focused on researching the benefits and challenges of AR for learning and training of students and employees as part of funded projects. (n = 12)

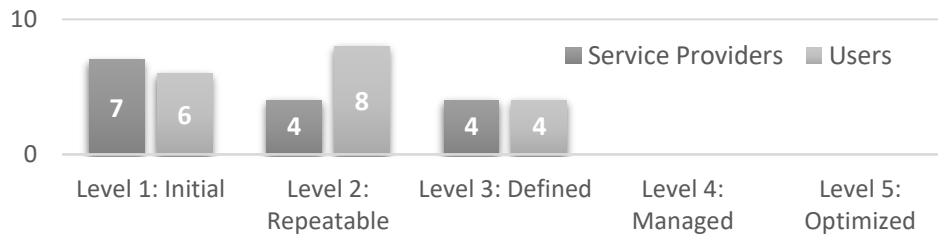


Fig. 1: Maturity levels of applications used in practice by service providers and users who took part in our interview study (n = 33). Levels are based on capability maturity models [Ma10, Pa93].

<b>Challenges regarding...</b>	<b>abs. #</b>	<b>rel. %</b>
<b>Adoption and Acceptance</b>	147	43%
Openness towards technology	62	42%
Recognition of added value	26	42%
Frustration of certain groups	21	34%
Fears about adopting	15	24%
Usability	46	31%
Low level of comfort	36	78%
Getting along with the system	6	13%
Resilience of users	4	9%
Integration in processes	29	20%
Exclusion of certain communities	10	7%
<b>Finance, Training, Infrastructure</b>	90	27%
Financing	30	33%
Training of employees	19	21%
Availability of hardware	13	14%
Internet coverage	12	13%
Industry sector specific issues	10	11%
Provision of 3D data	6	7%
<b>Hardware and Software</b>	45	13%
Limited hardware	20	44%
Realistic representation	12	27%
Limited software	9	20%
Complex setup	4	9%
<b>Uncertainty</b>	35	10%
Future impact possibilities	16	46%
Unknown area of expertise	14	40%
Unknown risks	5	14%
<b>Responsibility and Security</b>	22	6%
Responsibilities	11	50%
Security	4	18%
Ethical guidelines	4	18%
Privacy	3	14%

Tab. 1: Absolute numbers and relative frequencies of mentions for each group of codes and their respective sub-codes. The percentages always add up to 100% within each group.

### 3.2 Study Results

In this section we report on the dissemination, implementation in practice and challenges of AR for learning and training. After an overview over the current states of implementation, we present challenges that most likely have led to these states. In what follows, we focus mainly on the results gained from interviews with service providers and users (see above), as researchers often reported on prototypes rather than implemented

solutions and would hence bias the results. Among the challenges reported by our interviewees, we identified five different groups of challenges (see Tab. 1). These different groups will be described in order of their total amount of mentions and supplemented with their relative frequencies. Corresponding quotes were translated from German to English.

### **Implementation and Maturity Level of AR for Learning and Training in Practice**

The interviewees (only service providers and users) mentioned 64 unique examples of applying AR for learning and training in practice. Half of these were only mentioned briefly or in a state of ideas rather than actually being implemented. To understand the current state of applications, we focused on examples described in more detail and analyzed their maturity. Based on capability maturity models in general [Pa93] and in the context of e-learning [Ma10], we assigned them to levels of maturity. As Fig. 1 shows, we had enough information on a total of 33 application examples from service providers and users. The figure also shows that the majority of projects remain on initial steps with only prototypical use (if at all, level 1) or very initial usage by only a few users (level 2). Only eight projects fell into the category of defined applications that are systematically used for a small purpose (level 3). To our surprise, we did not find any reports on applications that reached a level above level 3 showing that their current state is still quite immature.

### **Challenges: Adoption and Acceptance**

When asked about challenges, the interviewees most frequently mentioned acceptance and adoption (about 43% of total mentions). On a closer look, interviewees therein described the openness towards the new technology (42%) and its usability (31%) as challenges. Others mentioned the integration of AR into processes (20%) as well as the exclusion of specific communities (e.g. visually impaired people, 7%).

Regarding acceptance, participants stressed the need to ascertain and communicate the added value of AR (42%): *“So, you need to know, to spark everyone in the company, [...] we created a demo that presented all the different things that could be possible with this technology.”* (user). Interviewees also reported on people having issues with or fears about adopting AR (24%): *“If people have no idea about the interactions, these people won’t put it on because of fear. They don’t want to embarrass themselves.”* (service provider).

Regarding the usability of AR, participants stated that they experienced a rather low level of comfort interacting with the hardware (78%) and issues getting along with provided software systems (13%). Regarding the comfort, we were told that *“a general challenge is a) when doing that for a long time and having those glasses on for like over half an hour or something, then you will notice a difference. One needs some time to calm down [afterwards]. The second thing is just the temperature. The temperature underneath those glasses.”* (service provider). People also explained that when talking *“[...] with firefighters, students and also participants, who tried those things, it still seems too new. It is too complex for now. It is too difficult to handle. Especially, we realize it when people have – I’d say – no experience with computer games or something like that. They basically have a bunch of problems interacting with the system.”* (researcher). In this context, it

was sometimes assumed that young people might adopt AR easier because of their affinity towards technology: *“I believe that young people like our trainees about 22 or 23 years old and older are quite good with technology. They are fairly open to play with it, like just using it. That’s why I don’t really see a problem there.”* (service provider).

### **Challenges: Finance, Training, Infrastructure**

Regarding finances, training and infrastructure as prerequisites for using AR, participants pointed at issues about their ability to roll out and use this technology at all (about 27% of all mentions). More specifically, a lack of knowledge on training of employees was mentioned often (21%). This was reported for both employees and trainers: *“In order to have trainees acquire those competences, the trainers need to be trained. And that’s what we are doing right now. But how do you teach [them] to impart AR?”* (service provider). In addition, some participants reported that they could not provide essential prerequisites like Wi-Fi access for AR devices (13%): *“So, when I look at the Wi-Fi connection at our project partners, well, then I guess this might take like ten, fifteen years”* (researcher). Also, the scarce availability of AR-devices was mentioned (14%). Some users stated that they only possessed a low number of devices that still take up a lot of work: *“Because I still have one device, I still have to/still need to ensure that it is always up to date. Eventually, you need to service it at the customer by the guys in our IT department, and somebody has to take care of that. And for some people this is just an ‘Oh, gosh. Not this as well on top.’-reaction.”* (service provider). Others told us that this was likely to improve, as companies seem to be more willing to offer funding for AR (33%). Although the acquisition cost has been decreasing, making devices more affordable, *“[it] is still relatively high”* and does not incorporate additional costs: *“I not only need to get the hardware. Most likely that is the smaller problem, but rather I need to change my way of working and everything that has to do with changing habits”* (service provider).

### **Challenges: Hardware and Software**

Regarding the capabilities of AR hard- and software (13% of total mentions), many participants pointed at the limited features and sensors of the hardware (44%). These remarks were specifically aimed at current head-mounted displays (HMD): *“I would just say that technically speaking, we are currently reaching the limits of the HoloLens regarding its abilities and features: Voice control, speech output, gesture control, object recognition.”* (user). People also mentioned other limitations such as resolution, GPS accuracy, field of view (FOV), cable management or gesture recognition. For many, this made the hardware seem to be insufficient and not close enough to be an authentic integration of reality and virtuality (27%), resulting in issues similar to the uncanny valley effect [SN07]: *“[We need something] Where you are training dialogue options and where you can see each other, not some other avatar in front of you. Although it might also be talking to you, it looks way different than the actual person behind the voice. [...] And when it is all about having a serious talk with someone to clarify a difficult situation where a conflict or something of that kind is going on [...] then people want to see gestures and facial expressions”* (user).

Some interviewees also remarked that the setup was too complex (9%) reducing the utility for spontaneous sessions: *“Now, looking at the Microsoft HoloLens for example, there is a complete Windows-PC on it, which means that I have to do more or less my complete Windows-Login and boot and start and open the app and all of that until I eventually can use my e-learning software. These are all very time-consuming steps that need to be shortened, so that a user is able to reach the contents it’s all about way faster.”* (service provider). In addition, many existing applications are mostly prototypes or work in progress, excluding them from any safety critical system (20%): *“[The tool] was about pacemakers and interacting with them by using an app. [...] And two years ago, there was a case that a hacker group thought they are going to do something fun and were sending everybody using these pacemakers a pushup notification [...] and told them ‘Hey, if you are not paying, we will turn your pacemaker off.’”* (user).

#### **Challenges: Uncertainty**

Uncertainties around AR amplify concerns on its added value (10% of total mentions). While there is a number of scenarios on how AR may impact the future, it is unclear to participants how this technology will affect their work (46%): *„So, this is not really so heady that one can actually say something. It still doesn’t grow on me why I should put on those helmets. Same for Virtual Reality. It is possible to do it nowadays [...]. But regarding learning situations, it is not tested enough and still in the fledgling stage.”* (researcher). As a consequence, participants are concerned about being exposed to a new and hence unknown situation (40%) and thus commented on probably facing unforeseen risks (14%): *„Mainly, it is this fear of being overwhelmed by the technology. I believe this is the root of it and along with it additional overhead that they later on can’t compensate for with these new possibilities.”* (service provider). These factors make it especially difficult for service providers to convey its added value regarding learning and training.

#### **Challenges: Responsibility and Security**

Regarding the usage of AR in practice (as opposed to trials and prototypes), additional concerns were raised (about 6% of total mentions). Some interviewees felt that legal responsibilities of using AR devices are quite uncertain, especially if users are being guided by AR (50%): *“Who is going to be responsible when a user is really just doing what he is told by those glasses and something happens while doing that? There, I see a huge problem regarding legal stuff.”* (user). In case of system failure resulting from malfunctions, we were told that laws are only starting to be discussed, adapted or implemented, if at all. Adding to this, some mentioned that ethical guidelines are still missing (18%): *“When I for example simulate a machine like a saw or something and put a piece of wood in there, can I cut off my hand with it? Or how will it work when I do something wrong? [...] If the engineers that develop this today have no idea about learning, this is sadly quite a high responsibility that [designers and providers] currently have to carry.”* (researcher). Interviewees also mentioned privacy problems through recordings and data analysis of employees for optimization reasons as unsolved (14%). Lastly, the reveal of corporate secrets and other sensitive information is a challenge

mentioned by participants (18%): „*So, the water supplier didn't want to hand over the plans for the drinking water pipelines because of information security and safety reasons. Because they think that at two or three locations, when you know them, you can bring down a whole city with poison. Or – I don't know –, maybe use explosives to destroy something. So, working with actual data is going to be difficult in some cases.*” (researcher).

## 4 Discussion

The insights provided by the interviewees show that despite its potentials – which initially caused them to engage with AR for learning and training – there are only a few applications of AR for learning and training outside the status of research projects. In our sample, many of these applications were in rather immature stages. This is also reflected in challenges related to insufficient hard- and software or the value added. The latter challenge in particular suggests that there is a discrepancy between expectations or assumptions and ways to implement these potentials in practice.

Some of the challenges may be overcome with systematic measures or over time, while others depend on additional factors. In the interviews, we could identify a tension between the questions about value added and the cost and availability of devices: As long as there is no clear value added, financial investments are low and hence availability will be slim. However, the other way around, if there are no devices available, it will be hard to investigate value added beyond small installations and prototypes. This, in turn, may also explain the low overall maturity of projects reported on by our interviewees (see Fig. 1). Additional uncertainties about developments around AR as a new technology amplify this. To break this tie, we need leading themes or best practices through which we can show how a systematic use of AR enhances learning and training success compared to other currently used means. We also need ways to analyze and communicate its potential value added and identify how to continually achieve it. Communicating good practice examples of AR (including a proof of their value added) as well as ways for training people to understand and use the technology and its potentials for their work could be viable steps towards a solution. They may help reduce reluctance and solve issues for adopters with little technological affinity by showing how benefits may outweigh existing concerns. In contrast to these challenges, problems such as lacking comfort when wearing an AR HMD, limited software and hardware support or problems in handling AR devices may diminish over time as AR continues to be researched, implemented and used. In a similar vein, issues of IT security can (in general) be solved by manufacturers of hardware and providers of software as soon as the demand increases and know-how is established.

From a different perspective, some of the challenges point at a lack of readiness for the application of AR in practice. Challenges such as how to train coaches or the lack of Wi-Fi connectivity suggest that many companies that could benefit from AR for learning and training may lack important prerequisites. If there are no coaches capable of teaching



employees on how to use AR, it will be difficult to roll out corresponding solutions. Likewise, a lack of Wi-Fi connectivity will directly translate into a lack of feasibility if digital tools are to support existing workflows. One may argue that companies that are already using AR may not have the resources to provide infrastructure or training for employees, but the fact that these issues were also mentioned by service providers, who offer AR based trainings, shows how important it is to focus on readiness. Complemented by insights regarding information security and data usage, these are issues to be solved in order to make users and services providers ready to use AR for learning and training.

There are also challenges concerning the responsibility and ethics of being guided by AR. Both are areas that currently also come up with technologies such as artificial intelligence. This shows how important it is to deal with them intensively and to come up with clear rules and norms. However, it is surprising that despite the importance and prevalence of these questions in German-speaking countries, they were mentioned less frequently than other challenges (6% total). This may indicate that we are currently too far away from implementing AR in practice, rendering these questions as unimportant for now.

Overall, our interview study shows that there is great interest in applying AR for learning and training. We could identify a large number of potential interviewees working in this area, and we received rich descriptions of ideas and applications. However, our study also clearly shows that besides its potentials, the implementations of AR in practice are in a rather immature stage, and that there are still several challenges that need to be overcome. Since some of these challenges are on a very basic level, this underpins the notion that there is still a lot of work to be done until AR becomes an established medium for learning and training in companies. The differences between expectations and questions or issues regarding added value, usability, adoption and others suggest that there is a need to find better ways of implementing potentials in practice. Because many currently available studies look at short-term (lab) studies, we need to show how AR unfolds long-term potentials and, thereby, justifies investments into hardware, software, infrastructure and training needed for its implementation in practice. Research may add to this by looking at and reporting on the benefits from long-term usage of AR for learning and training.

## **5 Limitations**

There are certain limitations that need to be recognized. First, the study builds on 44 volunteer interviewees from the areas of research, user organizations and service providers. They have been recruited from nearly 300 candidates and represent a good bandwidth of relevant actors. A study like this, however, is necessarily limited to a certain set of participants. Second, a study like this leaves little room for scrutinizing the statements and opinions we were told, leaving alone any chance for measurements or the like. However, we found many interviewees to mention similar aspects and challenges suggesting that these are present in many cases. In addition, the fact that very few interviewees mentioned established practices of AR for learning and training in practice

at all (cf. Fig. 1) suggests that there are only a few, if any, as it is very likely that we would have been told about successful cases first and intensively. Third, the study relies on information provided voluntarily by the interviewees without our knowledge of whether they have been withholding information. Despite these disadvantages that are inherent to interview studies, we are convinced that our work provides relevant insights into the use of AR for learning and training in practice. Interviews provide the opportunity to dive deep into experiences and challenges while providing good examples of what is happening.

The study presented here was conducted in late 2018, which on first sight may seem like it is slightly outdated. However, many devices such as the HoloLens are still state of the art and most issues and challenges found in the interviews are still valid to this date as we have been told by partners. Therefore, we consider the study to be very up-to-date and to present issues that need to be worked on in order to tap into the potential of AR for learning and training in practice. Since the interview study itself only focuses on companies in German-speaking countries, statements about the international situation of AR regarding learning and training in practice may be inappropriate. Regardless, in our work, we discuss challenges faced by a number of companies in an economically important country.

## 6 Conclusion

In our work, we present potentials of AR for learning and training based on current research and contrast these findings with results of an interview study conducted with 44 participants from German-speaking companies. While literature shows benefits for AR in learning and training, we asked participants about their experiences with AR. Based on these answers, we discovered that many AR applications for this purpose are still in an early stage and, hence, rather immature. Also, we were told about a number of challenges still being faced when implementing AR in companies. However, besides challenges like uncertainties regarding the value added, a lack of provided infrastructure, hard- and software limitations and the absence of know-how and legal norms, our study shows that AR is considered relevant for learning and training in practice. In order to further increase the maturity of applications and provide insights on the long-term values added, challenges and benefits need to be compared and systematically worked on, especially in long-term studies. This way, more companies could be willing to invest in implementing this technology and training users.

## Bibliography

- [Ab17] Abramovici, M. et al. 2017. Context-aware Maintenance Support for Augmented Reality Assistance and Synchronous Multi-user Collaboration. *Procedia CIRP*. 59, (2017), 18–22. DOI:<https://doi.org/10.1016/j.procir.2016.09.042>.
- [Az18] Azimi, E. et al. 2018. Evaluation of Optical See-Through Head-Mounted Displays in Training for Critical Care and Trauma. 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (Mar. 2018), 1–9.

- [Az97] Azuma, R.T. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*. 6, 4 (Aug. 1997), 355–385. DOI:<https://doi.org/10.1162/pres.1997.6.4.355>.
- [Bo14] Bower, M. et al. 2014. Augmented Reality in education – cases, places and potentials. *Educational Media International*. 51, 1 (Jan. 2014), 1–15. DOI:<https://doi.org/10.1080/09523987.2014.889400>.
- [Bü20] Büttner, S. et al. 2020. Augmented Reality Training for Industrial Assembly Work – Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training? *Proceedings of the 2020 ACM Conference on Human Factors in Computing Systems (CHI 2020)* (2020).
- [Bü17] Büttner, S. et al. 2017. The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach. *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (Island of Rhodes, Greece, 2017)*, 433–440.
- [Di13] Di Serio, Á. et al. 2013. Impact of an Augmented Reality System on Students’ Motivation for a Visual Art Course. *Comput. Educ.* 68, (Oct. 2013), 586–596. DOI:<https://doi.org/10.1016/j.compedu.2012.03.002>.
- [Du18] Duerr, M. et al. 2018. HCI meets Nursing Care - The application of Mixed Reality in basic Nursing Care Education. *Zukunft der Pflege: Tagungsband der 1. Clusterkonferenz 2018 - Innovative Technologien für die Pflege (Oldenburg, Jun. 2018)*, 100–105.
- [Fa16] Fakourfar, O. et al. 2016. Stabilized Annotations for Mobile Remote Assistance. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (New York, NY, USA, 2016)*, 1548–1560.
- [Fu12] Fujimoto, Y. et al. 2012. Relation Between Location of Information Displayed by Augmented Reality and User’s Memorization. *Proceedings of the 3rd Augmented Human International Conference (New York, NY, USA, 2012)*, 7:1–7:8.
- [Fu17] Funk, M. et al. 2017. Working with Augmented Reality?: A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. *Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (New York, NY, USA, 2017)*, 222–229.
- [Ho17] Hochberg, J. et al. 2017. Gestaltung und Erforschung eines Mixed-Reality-Lernsystems. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*. 28, (2017), 140–146.
- [Ki75] Kirkpatrick, D.L. 1975. *Evaluating training programs*. McGraw-Hill Education.
- [Ko18] Kopetz, J.P. et al. 2018. Eignung von Datenbrillen zur Unterstützung von Pflegekräften in der Ausbildung. *Zeitschrift für Arbeitswissenschaft*. 72, 1 (2018), 13–22.
- [Li17] Linn, C. et al. 2017. Virtual Remote Inspection - A new Concept for Virtual Reality enhanced real-time Maintenance. (Nov. 2017).
- [Ma10] Marshall, S. 2010. A Quality Framework for Continuous Improvement of e-Learning: The e-Learning Maturity Model. *Journal of Distance Education*. 24, 1 (2010), 143–166.

- [MK94] Milgram, P. and Kishino, F. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*. 77, 12 (1994), 1321–1329.
- [Mo17] Mourtzis, D. et al. 2017. Augmented Reality Application to Support Remote Maintenance as a Service in the Robotics Industry. *Procedia CIRP*. 63, (2017), 46–51. DOI:<https://doi.org/10.1016/j.procir.2017.03.154>.
- [Ni17] Niemöller, C. et al. 2017. Sind Smart Glasses die Zukunft der Digitalisierung von Arbeitsprozessen? Explorative Fallstudien zukünftiger Einsatzszenarien in der Logistik. (2017).
- [Pa18] Pai, Y.S. et al. 2018. UbiTrain: Leveraging the Physical and Virtual Environment for Ubiquitous Sports Training. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore, Oct. 2018), 202–206.
- [Pa93] Paulk, M.C. et al. 1993. Capability maturity model, version 1.1. *IEEE Software*. 10, 4 (Jul. 1993), 18–27. DOI:<https://doi.org/10.1109/52.219617>.
- [Ph96] Phillips, J.J. 1996. Measuring ROI: The Fifth Level of Evaluation. *Technical & Skills Training*. (Apr. 1996), 10–13.
- [RS19] Radu, I. and Schneider, B. 2019. What Can We Learn from Augmented Reality (AR)? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk, May 2019), 1–12.
- [Sa16] Santos, M.E.C. et al. 2016. Augmented reality as multimedia: the case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*. 11, 1 (Dec. 2016), 4. DOI:<https://doi.org/10.1186/s41039-016-0028-2>.
- [SN07] Seyama, J. and Nagayama, R.S. 2007. The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments*. 16, 4 (Aug. 2007), 337–351. DOI:<https://doi.org/10.1162/pres.16.4.337>.
- [Si17] Siebert, J.N. et al. 2017. Adherence to AHA Guidelines When Adapted for Augmented Reality Glasses for Assisted Pediatric Cardiopulmonary Resuscitation: A Randomized Controlled Trial. *Journal of Medical Internet Research*. 19, 5 (May 2017), e183. DOI:<https://doi.org/10.2196/jmir.7379>.
- [Ul16] Ullrich, C. et al. 2016. APPSist Statusbericht: Realisierung einer Plattform für Assistenz- und Wissensdienste für die Industrie 4.0. *DeLFI Workshops* (2016), 174–180.
- [We18] Werrlich, S. et al. 2018. Assembly training: comparing the effects of head-mounted displays and face-to-face training. *International Conference on Virtual, Augmented and Mixed Reality* (2018), 462–476.
- [Wi19] Wilschut, E.S. et al. 2019. Evaluating learning approaches for product assembly: using chunking of instructions, spatial augmented reality and display based work instructions. *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (2019), 376–381.
- [Zh14] Zhu, E. et al. 2014. Augmented reality in healthcare education: an integrative review. *PeerJ*. 2, (2014), 469.

# LEAP: Learnings beim Einsatz von Augmented Reality in der praktischen Berufsausbildung

## Lernhaltigkeit mit Augmented Reality angereicherter Trainings

Jens Hofmann<sup>1</sup>

**Abstract:** Der Einsatz von Augmented Reality (AR) steckt, besonders in der beruflichen Bildung, noch in den Kinderschuhen. Dies resultiert insbesondere aus den fehlenden Erfahrungen im Umgang mit der Technologie und verhindert dabei gleichzeitig die einfache Generierung maßgeschneiderter sowie lernhaltiger Inhalte. In LEAP steht die Anwendung sog. AR-basierter Remote-Trainings, in der praktischen Ausbildung von Chemikanten, im Mittelpunkt. Die untersuchten Szenarien führten zu einer verstärkten Reflexion der Lernenden, insbesondere durch den Einsatz unterschiedlicher Kommunikations- und Kollaborationsformen. Dies deutet auf ein hohes, integratives Potential von AR in bestehende Lern- und Arbeitsprozesse hin.

**Keywords:** Augmented Reality, Remote Training, berufliche Bildung, Chemikant

## 1 Einleitung

Die Anwendung von Augmented Reality, in Arbeitsprozessen, steckt derzeit noch in den Anfängen. Erste Erfahrungen sind vielversprechend. Hinzu kommt, dass die zunehmenden Bedarfe nach digital gestützten oder digital basierten Lösungen, infolge der weltweiten Pandemie, ab Februar 2020, entsprechende Entwicklungen beschleunigen.

Unter Augmented Reality (AR) wird dabei "... a medium in which digital information is overlaid on the physical world that is in both spatial and temporal registration with the physical world and that is interactive in real time" [Cr13] verstanden. Die dreidimensionale Verankerung virtueller Informationen im Raum, mittels 3D Scan in Echtzeit, mit der entsprechenden Hardware, führt zur Mixed Reality (MR). Dies vereint die Vorteile der Virtual Reality (VR), durch die Nutzung dreidimensionaler Objekte, sowie von Augmented Reality, durch die Einbindung dieser Objekte in die reale und sichtbare Umgebung. Diese Anreicherung der Realität mit entsprechenden virtuellen Objekten ist wichtig beim Einsatz in realen Arbeitsprozessen.

---

<sup>1</sup> Sächsische Bildungsgesellschaft für Umweltschutz und Chemieberufe Dresden mbH (SBG Dresden),  
Projektmanager Augmented Reality, Gutenbergstraße 6, 01307 Dresden, j.hofmann@sbgdd.de

Die Tätigkeit in industriellen Arbeitsumgebungen ist u.a. geprägt durch Einschränkungen höhere Geräuschpegel als auch die Notwendigkeit beide Hände für Arbeiten zu nutzen. Aus diesem Grund sind AR-Anwendungen mit Smart glasses oder Datenbrillen der Nutzung auf Smartphones, Tablets und Laptops zu bevorzugen.

AR bietet sich bei risikoarmen Aufgaben, wie Wartungs- und Montagearbeiten, an. In der praktischen Ausbildung, in einer Trainingsumgebung, können darüber hinaus Handlungsmuster von Auszubildenden eingeübt werden, die dann auf den jeweiligen Maschinenpark ihrer Ausbildungsunternehmen übertragen werden können. Bisherige Studien, zur Anwendung von AR, fokussierten sich besonders auf den technischen Service [Ga11] und teilweise der beruflichen Bildung [LH20]. Weitere Veröffentlichungen [Ga11] [GA19] zeigen, dass der Einsatz von AR/MR meist zu einem motivierteren und effektiveren Lernen führt.

Der Anwendung von AR/MR hat Vorteile bei der Vermittlung beruflicher Handlungskompetenz. In Kombination mit passenden Kommunikations- und Kollaborationsformen entstehen neue Kommunikationsräume für Entwicklung, Test, Evaluation und Transfer lernhaltiger und mit AR angereicherter Lehr- und Lernszenarien.

In LEAP wird der Einsatz von AR bei der Steuerung von Prozessen, in einer chemischen Pilotanlage, untersucht. Im Mittelpunkt steht dabei das Berufsbild des/der Chemikanten/-in, im 1. Lehrjahr.

## 2 Theorie

Augmented Reality / Mixed Reality fördert einerseits ein experimentelles und selbstgesteuertes Lernen und andererseits ein angeleitetes, handlungsorientiertes Lernen. Gerade das Lernen unter Anleitung, z.B. eines betrieblichen Ausbilders, zeichnet die berufliche Erstausbildung aus. Ein theoretischer Bezugsrahmen hierfür ist die sog. Cognitive Apprenticeship Theory [CBN89]. Der konstruktivistische Ansatz steht dabei in enger Verbindung zum situierten Lernen.

Der Cognitive Apprenticeship ist durch die aufeinanderfolgenden Phasen: Vorführen - unterstützte Vorführung – Artikulation und Reflexion des Gelernten – Übertragung des Gelernten auf neue, aber ähnliche Aufgaben gekennzeichnet. Er bildet damit die Realität in der Ausbildung (Auszubildender-Meister/Ausbilder-Verhältnis) ab, bei der das Wissen in einem zyklischen Verhältnis zur Praxis steht. Dieses Wissen wiederum beeinflusst das Handeln, was wieder zu neuem Wissen führt. Die Wechselbeziehung von Wissen und Handlungswissen bzw. dessen Reflexion ist ein Merkmal für das Lernen im Prozess der Arbeit. Der Cognitive Apprenticeship-Ansatz lässt sich damit bei der Nutzung von AR/MR in der praktischen Ausbildung bei Chemikanten (1. Lehrjahr) anwenden, da bestehende, sequenzielle Arbeitsschritte (audio-)visuell und synchron begleitet werden, um das jeweilige Lernziel, inhaltlich und zeitlich, zu erreichen.

### 3 Anwendungsszenarios

Die Verwendung von AR/MR in der praktischen Ausbildung, im Bereich der Chemie, ist überwiegend Neuland, aber geboten. Treiber der Entwicklung ist die Novellierung des Berufsbildes Chemikant/-in, im Jahr 2018, durch die Einführung der Wahlqualifikation „Digitalisierung und vernetzte Produktion“. Dies inkludiert dabei auch den Umgang mit Technologien wie AR. Die SBG Dresden als überbetriebliche Bildungseinrichtung, mit dem Fokus auf dem praktischen Teil der Ausbildung, ist bestrebt die Anforderungen der Ausbildungsordnungen und der Ausbildungsunternehmen entsprechend abzubilden.

Aufgrund des Fehlens bestehender Erfahrungswerte beim Umgang mit AR, in der praktischen Ausbildung, erfolgte die Pilotierung bei ausgewählten Arbeitsaufgaben im zweiwöchigen „Verfahrenstechnischen Praktikum I“, für Auszubildende des 1. Lehrjahres. Hierbei kam das sog. **Remote-Training**, für den Erwerb von Handlungswissen zum Einsatz. Die Ergebnisse wurden anschließend formativ evaluiert.

#### **Szenario I: Remote-training (Ausbilder-Auszubildender)**

Bei diesem Szenario erhält der Auszubildende, welcher die AR/MR Brille (Microsoft HoloLens 1) trägt und die Microsoft-App „Remote Assist“ auf der Brille geöffnet hat, in Echtzeit audio-visuelle Anweisungen vom Ausbilder. Dieser ist über Microsoft Teams zugeschaltet und sieht über die Kameras der Datenbrille, was der Auszubildende sieht. Damit sind expertenbasierte AR-Szenarien einfach umsetzbar, da durch die App-Nutzung kein Programmieraufwand entsteht, um Hologramme zu erzeugen.

Für die Pilotierung wurden vier Chemikanten sowie eine Chemikantin ausgewählt. Diese waren bereits drei Tage lang in der chemischen Pilotanlage („Technikum“) der SBG, während ihres dreiwöchigen, sog. verfahrenstechnischen Praktikums. Nach einer kurzen Einführung, im Umgang mit der Microsoft HoloLens 1 sowie eine fragebogenbezogene Erfassung der Motivation und des bereits bestehenden, möglichen Vorwissens, um festzustellen ob die jeweilige Aufgabe nicht bereits vorher ausgeführt hatten, begann die AR/MR gestützte Umsetzung der Aufgaben (siehe Tabelle 1).

Der Ausbilder wählte anschließend über seinen Laptop den Nutzer an, der die HoloLens trug. Nach der Annahme des eingehenden Anrufs, in der Remote-Assist App durch den Auszubildenden, war es dem Ausbilder möglich audio-visuelle Anweisungen in Echtzeit, im aufgebauten Videostream, zu geben. Er machte mit Hologrammen (u.a. in Pfeilform) aufmerksam auf Ventile u.a., die der Auszubildende bei der Ausführung der Aufgaben beachten sollte. Zur besseren Kommunikation bei gleichzeitiger Verringerung des Einflusses von Umgebungsgeräuschen in der Anlage trug der jeweilige Auszubildende Kopfhörer, die mit der Audiobuchse der HoloLens-Brille verbunden waren. Damit war es möglich den räumlich entfernten Ausbilder besser zu verstehen.

Das AR basierte Remote-training ermöglichte die Ausführung der jeweiligen Aufgabe in Echtzeit zu steuern. Fehler waren sofort korrigierbar, der Lernerfolg war jeweils direkt evaluierbar.

Die durchgeführten Pilotierungen wurden vom Ausbilder, durch die jeweilige, nachträgliche Einschätzung mit einem kurzen Fragebogen, hinsichtlich der Operationalisierbarkeit der Arbeitsschritte (einfach/bedingt/nicht) von AR/MR in der praktischen Ausbildung, ausgewertet. Es zeigt sich, dass die gewählten mit AR/MR angereicherten Lehr- und Lernszenarien, hinsichtlich vermittelter Inhalte und kommunikativer Struktur, lernhaltig sind. Dies stellte sich insbesondere in der effektiven Vermittlung von Handlungswissen sowie der besseren Reflektion der ausgeführten Tätigkeiten dar.

Test	Aufgabe	Niveau	Dauer	Evaluation (multiple und aufeinanderfolgende Arbeitsschritte)	Ausbilder- oder Auszubildendenorientiertes Szenario
1	Pumpen einer Salzlösung von Behälter B2 zu B4	Anfänger	7 min	Einfach durchzuführen	Ausbilder - Auszubildender
2	Pumpen einer Salzlösung von Behälter B2 zu B4 und Wechsel Filter sowie Pumpenschlauch	Anfänger	27 min	Einfach durchzuführen	Übergang von Auszubildenden (aus Test 1) – Auszubildenden zu Ausbilder- Auszubildender
3	Probenahme der Salzlösung (Salzgehalt- und Dichtebestimmung mit Aräometer)	Anfänger	16 min	Einfach durchzuführen	Ausbilder - Auszubildender
4	Filtern einer Salzlösung: Wechsel Filter	Anfänger	5 min	Einfach durchzuführen	Ausbilder - Auszubildender
5	Filtern einer Salzlösung: Wechsel Filter	Anfänger	6 min	Einfach durchzuführen	Ausbilder - Auszubildender

Tab. 1: Übersicht durchgeführte Testsessions (Chemikant, 1. Lehrjahr)

### Szenario II: Remote Training (Auszubildender-Auszubildender)

Neben der ausbilderunterstützten Ausführung der jeweiligen Aufgabe wurde auch ein Wechsel der Kommunikations- und Kollaborationsform untersucht. Die Vermittlung des Handlungswissens durch einen zuvor vom Ausbilder geschulten Auszubildenden wurden für das Remote-Training erprobt („Test 2“). Es zeigte sich, dass diese Art des AR gestützten Peer-Learnings einfach umsetzbar ist. Der erfahrene Auszubildende nahm vor dem Laptop des Ausbilders Platz und gab dem Auszubildenden, mit der HoloLens 1, audio-visuelle Anweisungen in Echtzeit. Die Dauer ist mit der des Ausbilder-Auszubildenden Szenarios vergleichbar. Damit konnte auch der Wissenstand des



vermittelten Auszubildenden überprüft werden. Wie in Szenario I wurde Szenario II vom Ausbilder, durch die nachträgliche Einschätzung mit einem kurzen Fragebogen, hinsichtlich der Operationalisierbarkeit von AR/MR (einfach/bedingt/nicht) in der praktischen Ausbildung, ausgewertet.



Abb. 1: Remote-Training Session (Test 4)

#### 4 Diskussion

Der Einsatz von AR/MR in der praktischen Ausbildung ist einfach möglich, da ein gesonderter Programmieraufwand für eine sich auf die berufliche Aus- und Weiterbildung fokussierende Einrichtungen eher problematisch ist. Remote-Trainings bieten sich bei der Vermittlung von Handlungswissen, z.B. verstärkt im 1. Lehrjahr, an. Damit ist es möglich Anlernprozesse zu unterstützen, wenn gleichzeitig ein Ausbilder oder eine erfahrene Person zur Verfügung steht. Die durchgeführten Pilotierungsszenarien mit einer kleinen Stichprobe verdeutlichen, dass eine AR/MR basiertes Remote-Training besonders bei eher komplexen bzw. aufeinander aufbauenden Arbeitsschritten zum Einsatz kommen kann. Wichtig ist, dass der Erwerb von Handlungswissens gerahmt werden muss, durch den vorangeschalteten Erwerb von Fachwissen, zur Funktionsweise der augmentierten, technischen Anlagen und -teile.

Die Änderung der Kommunikations- und Kollaborationsform, hin zu einem Peer Learning Szenario, ist lernhaltig. Eine gesteigerte Motivation der Lernenden wird vermutet u.a.

infolge der gemeinsamen, digitalgestützten und teilweise zeiteffektiveren Zusammenarbeit. Damit können bestehendes Wissen und Fähigkeiten beim instruierenden Auszubildenden evaluiert werden sowie die Effektivität der Aufgabenausführung des instruierten Auszubildenden.

## 5 Schlussfolgerungen

Die Anwendung immersiver Technologien in der praktischen Ausbildung gewerblich-technischer Berufe ist realisierbar, wenn diese einfach umzusetzen sind und insbesondere die Kommunikation und Kollaboration in den Mittelpunkt stellen. Ein besonderer Fokus sollte dabei zukünftig auf der Lernhaltigkeit von AR/MR basierten Szenarien, sowohl in der Tiefe (z.B. komplexere Aufgabe, weitere Kommunikations- und Kollaborationsszenarien) als auch Breite (weitere Industrien) liegen. Maßgeschneiderte Anwendungsszenarien mit klarem didaktischem Mehrwert beschleunigen die Einbindung in (a-)synchrone und (in-) direkte Szenarien. Der didaktische Mehrwert aufgezeichneter Remote-Training Sessions, in Form von Lernvideos, ist ergänzend zu untersuchen.

## Literaturverzeichnis

- [Ga11] Gavish, N. et. al.: Design guidelines for the development of virtual reality and augmented reality training systems for maintenance and assembly tasks. In BIO Web of Conferences 1: International, 2011.
- [GA19] Garzón, J.; Acevedo, J.: Meta-analysis of the impact of augmented reality on students' learning gains. Educational Research Review, 27, S. 244-260, 2019.
- [CBN89] Collings, A.; Braun, J. S. and Newman, S. E.: Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing and Mathematics. In: Resnick, L.B. (Hrsg.) Knowing, Learning and Instruction. Essays in Honour of Robert Glaser. Hillsdale, NJ., S. 453-494, 1989.
- [Cr13] Craig, Alan B.: Understanding Augmented Reality. Concepts and Applications. Elsevier, Waltham, 20, 2013.
- [LH20] Lester, S.; Hofmann, J.: Some pedagogical observations on using augmented reality in a vocational practicum. British Journal of Educational Technology, Vol 51, No 3, S. 645-656, 2020.

## Comparing Virtual Reality and Screen-based Training Simulations in Terms of Learning and Recalling Declarative Knowledge

Leonard Meyer<sup>1</sup>, Thies Pfeiffer<sup>2</sup>

**Abstract:** This paper discusses how much the more realistic user interaction in a life-sized fully immersive VR Training is a benefit for acquiring declarative knowledge compared to the same training via a screen-based first-person application. Two groups performed a nursing training scenario in immersive VR and on a tablet. A third group learned the necessary steps using a classic text-picture-manual (TP group). Afterwards all three groups had to perform a recall test with repeated measurement (one week). The results showed no significant differences between VR training and tablet training. In the first test shortly after completion of the training both training simulation conditions were worse than the TP group. In the long-term test, however, the knowledge loss of the TP group was significantly higher than that of the two simulation groups. Ultimately, VR training in this study design proved to be as efficient as training on a tablet for declarative knowledge acquisition. Nevertheless, it is possible that acquired procedural knowledge distinguishes VR training from the screen-based application.

**Keywords:** Virtual Reality Training; Screen Based Training; Declarative knowledge acquiring; Health Care Scenario

### 1 Introduction

In recent years, virtual reality has found its way into the global consumer market. Commercial VR headsets are widely used for entertainment purposes whose underlying applications are exploited by developers in all directions of new interaction possibilities. VR technologies allow the user to see and interact with virtual environments from a first person fully immersive view. They elevate virtual worlds, as they have been known for decades from simulations, video games and videos, to a new level of user experience. But this modern technology does not have to serve only for entertainment, there is a high potential of new concepts for digital learning media, of which various implementations have already been examined for effectiveness in education and teaching in the past years.

VR offers a wide range of features that could be useful for teaching. Any imaginable 3D environment can be displayed independent of the real available space. Users can interact freely in virtual space and can receive direct audio, visual and even haptic feedback. On the other hand, various possibilities of virtual reality are also available in other media such as videos or even closer in screen-based simulations. The sticking point that distinguishes

---

<sup>1</sup> Hochschule Emden/Leer, Abteilung E+I, Constantiaplatz 4, 26723 Emden, leonard.meyer@hs-emden-leer.de,

<sup>2</sup> Hochschule Emden/Leer, Abteilung E+I, Constantiaplatz 4, 26723 Emden, thies.pfeiffer@hs-emden-leer.de

the different media is the way of learning. While videos are passive learning objects and therefore only allow displaying information, simulations of all kinds allow an active interaction with the environment. What this interaction looks like can make a difference in learning success, which makes it part of a wide area in educational research considering different types of learning. In general, the concept of distinct learning types is very controversial, with several models, such as the visual-auditory-kinaesthetic learning styles model [BMS88]. Still there are differences in the way of learning and in the success of different teaching methods, but these differences are difficult to integrate into clear structures. Nevertheless, VR offers a wider range of interaction (and thus information) channels compared to screen-based simulations, especially on realistic training movements in life-sized environments. Therefore a decisive factor, which should positively distinguish the learning success in VR simulations from other methods, is the self-performing of a training or a procedure by the user, even if done virtually with the limitations of the respective system (lack of haptic feedback, controller-based input or limited walking space).

This advantage seems obvious for acquiring procedural knowledge (learning how) and has been proven in a couple of studies in the last years. For example, in a study from 2018, VR training was compared to non-VR training in the field of surgical education [Xi20]. 16 doctoral students of surgery were divided into two training groups. The task to learn was the insertion of pedicle screws into the spine, which was subsequently evaluated on human cadavers. The subjects of the VR group performed a simulation of this procedure during training, the non-VR group received a classical introduction by the chief surgeon. In terms of success rate and precision as well as speed, the subjects of the VR group were significantly better in the test.

Beside this it is also a question of whether the acquisition of declarative knowledge (factual knowledge; learning what) can also be improved by running a training simulation in VR. Declarative knowledge as such can be taught interactively in a screen-based simulation as well as in a VR simulation. Therefore, the question is: Does the fact of completing a training in VR lead to a greater (declarative) learning success than completing the same training on a screen? Since the interaction possibilities in VR are much closer to reality, this could have a positive influence on the learning success, which would not be present in screen-based training. Furthermore, it is important to measure the long-term success: How much have learners forgotten again after a certain time?

Thus a study was designed which disregards the advantages of a VR training simulation in knowledge retrieval but not in knowledge acquisition and makes a comparison to a training technique without these extended possibilities. More precisely, the study deliberately avoided measuring procedural knowledge, where various possible advantages are obvious in VR training, especially with regard to later execution in reality.

## 2 Materials and Methods

### 2.1 Participants

The experimental sample included 45 participants, 21 women and 24 men with an average age of 27 (min 19, max 41) recruited from Bielefeld University (Germany). Nearly all participants were students, most of whom came from computer science, psychology or philosophy. Attention was paid to an even distribution of participants in the test groups.

### 2.2 Experimental Design

A univariate design with two repeated measurements was applied. The considered factor was the level of knowledge about the necessary steps in the conducted training.

The overall sample was divided into 3 equally sized subgroups. One of them conducted the training in VR, a second one completed a screen-based application (SB group). Since the study was thematically concerned with the acquisition of declarative knowledge, a third comparison group was tested to compare the learning success to pure memorization.

In this group, the subjects were to learn the necessary steps by heart using simple text and picture instructions without any further training (TP group). Each group got training specific standardized instructions in terms of handling. The two simulation groups ran through the scenario exactly once. The average time required by the subjects was then set as learning time for the TP group.

Measurements were taken in a pre-test before completing the respective training, in a subsequent post-test and in a second post-test one week later. As described by Ebbinghaus' forgetting curve, the strongest loss of uniquely learned knowledge takes place in the first days after learning and then flattens out [Eb85]. Thus, it could be assumed that after seven days a sufficiently high discrepancy of knowledge levels emerges.

### 2.3 Educational Content

In the context of the study, participants had to deal with a healthcare scenario. The main focus lied on learning the procedure of a medical infusion preparation, a scenario nursing students are supposed to learn. This training was developed as a fully immersive VR application as part of the Virtual Skills Lab Project [Pf18] and used for the VR test group.

The training included 23 separate steps to perform. The screen-based group got the same application in classical first-person point-and-click adventure layout, a video game genre that has been outlined because of its suitability for educational applications [Am01].

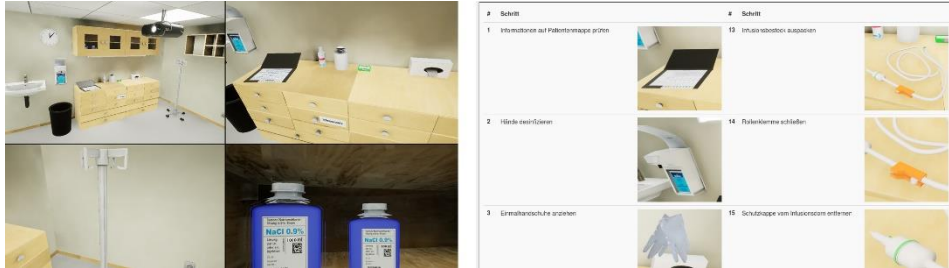


Fig. 1: Virtual lab in VR/screen-based app and instruction manual for third test group

Both virtual trainings were guided by the same instructions that were also used in the picture and text manual. The VR group received the instructions on a small widget on the left wrist, the SB group had the same widget in the upper left corner of the screen. Thus each group had the same training material at a different level of interaction. The TP group only had passive learning objects, the SB group had passive learning objects and an interactive environment and the VR group additionally had realistic behavior while interacting.

## 2.4 Knowledge Measurement

In general, a subjects knowledge level was defined as the set of known steps from the correct sequence both independently and dependent on the correct order. This way a participant who has forgotten the correct order but remembers the correct steps still knows more than someone who cannot remember anything.

To make the level of knowledge quantifiable, a digital flash card system had been developed and integrated in a web application. Objects used (bottle, disinfection cloth, ...) and activities performed (disinfecting, pricking, testing, taking, ...) were extracted from the individual steps of the infusion preparation and created as separate virtual "flash cards". The participants now had to build the tasks from the available objects and activities by combining them and putting them in an assumed correct order. In addition, distractor objects were added to the object pool, a rather obvious, one thematically roughly related but questionable and an absurd one: Bandages, scalpel and barbecue lighter. The latter served as a control item to eliminate unusable entities from the later data set. Since activities and objects were available in a data pool, they could be used as often as assumed for combining.

## 2.5 Calculating Distances Between Remembered and Original Procedures

If a lower case letter <x> is assigned to each object and an upper case letter <Y> to each activity, each created procedure can be described as a character string consisting of tuples of the type <xY> (e.g. *aDdEbCfN*; aD => "disinfect hands"). If the correct procedure is

now defined as *iJaDdAcIbDdGaDfJeJeKjLkFlFgEfBmHjKhCnC*, a second procedure can be checked for similarity. The more similar the character string to be checked, the smaller the number of errors in the created sequence.

The Damerau-Levenshtein distance (DL distance) was used to calculate the minimum number of operations required to change one character string into the other (insertions, deletions or substitutions of a single character, or transposition of two adjacent characters)[MVM09]. By using the DL distance, however, a calculation problem can occur. A respondent who submits the character string *iJ* as a whole sequence has 36 errors according to the calculation, because 18 steps of the correct sequence are missing and must be inserted. In comparison, a second participant submitting a sequence that corresponds exactly to the correct sequence, but backwards, would result in the same distance. Both operations have various errors, but in relation to learning success the second subject knows considerably more, since at least all objects and activities are present and have been correctly combined. Omitting or forgetting correct objects/activities must therefore be given more weight in the calculation.

Two distinct sets can be derived from the correct procedure: Correct objects and correct activities. To positively weight the remembering of objects (and likewise activities) to be used, this set can be compared to the set of objects actually used in the submitted process. The similarity of the two sets 'CorrectObjects' and 'CorrectObjectsUsed' should influence the error rate. Therefore Jaccard-Similarity is used. Applied to activities and objects, two percentage values result, which correspond to the percentage similarity to the correct set of objects or activities.

In total, the following formula results:

$$Error = \frac{DL D}{0.5 * (object.similarity + activity.similarity)}$$

If all objects and activities are used, the denominator of the formula results in 1 and only the DL distance describes the error rate. If there are less objects/activities to be used, the denominator becomes smaller, which has a negative effect on the error rate.

The first subject from the previous example (submitted sequence = *iJ*) now has an error rate of 65.44, while the second subject still has 36 errors. An additional advantage of the separate consideration of correct objects/activities is to enable distinct analyses.

## 2.6 Sequence Blur

To calculate the learning success, a concrete solution for calculating deviations is needed. However, this is difficult to define. The free combination of objects and activities (albeit from a fixed pool) leaves room for interpretation.

In general, the verb "to take" is too versatile to be integrated into a concrete solution string. In order not to withhold existing steps, "take" was available as a verb for participants, but

all created steps containing "take" were eliminated from the created character strings before the DL distance calculation.

Double interpretations were also handled. A step like "remove protection cap" has about the same meaning as "open protection cap".

## 2.7 User Experience Measurement

A difficult to use or distracting software can be a reason for a lower learning outcome. To exclude such cases, the User Experience Questionnaire (UEQ) was used to measure the usability of the training simulations [LHS08]. Participants had to fill out the questionnaire immediately after the first post-test (after the respective training).

## 2.8 Hardware

The VR group performed the training on a regular HTC Vive using the native Motion Controllers for interaction. The SB group used a Samsung Galaxy Tab S5e (10.5 inch). The TP group got the textual/picture instructions on a PC with a 24 inch monitor. All subjects performed the knowledge acquisition tests on a PC with a 24 inch screen.

# 3 Results

Informed written consent was obtained from each participant before the study. 5 out of 15 participants in the VR group stated to have already been using a VR application several times, 10 participants in the overall sample. The investigator observed that the handling of the interface for creating an infusion preparation sequence and the task definition was intuitively understandable for most of the subjects and for all after a brief explanation of the mechanics in the web application for the test runs. Training took place in a sound proofed room to ensure an undisturbed working atmosphere.

No participant was aware of the process of preparing an infusion before the start of the study and no one had integrated the control item "barbecue lighter" into the procedures created. Thus the entire data set was enabled for further analyses.

## 3.1 User Experience

Figure 2 highlights the mean values (columns) and variances ( $s^2$ ; whisker) of the measurement results. The scales lie between values from -3 (extremely bad) to 3 (extremely good). Statistically, however, values above  $\pm 2$  are considered very unusual [LHS08].



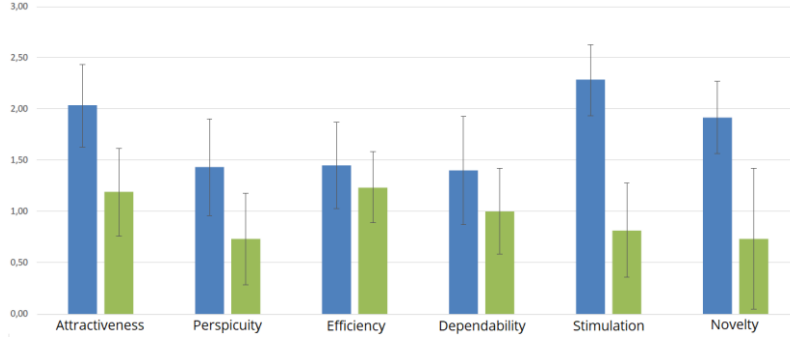


Fig. 2: UEQ Evaluation; blue = VR group, green = SB group

Overall the screen-based application received positive feedback. Attractiveness ( $M = 1,189, s^2 = .71$ ) and efficiency ( $M = 1,233, s^2 = .47$ ) are the highest values. Dependability ( $M = 1.0, s^2 = .68$ ) and stimulation ( $M = .817, s^2 = .83$ ) are also still in the good range. The values for perspicuity ( $M = .733, s^2 = 0.79$ ) and novelty ( $M = .733, s^2 = 1.68$ ) are in the upper average range.

In the VR version transparency ( $M = 1.433, s^2 = .86$ ), efficiency ( $M = 1.45, s^2 = .70$ ) and controllability ( $M = 1.4, s^2 = 1.10$ ) are in the good range with average values of 1.4. The originality follows higher ( $M = 1.917, s^2 = .49$ ). Attractiveness ( $M = 2.033, s^2 = .62$ ) and stimulation ( $M = 2.283, s^2 = .46$ ) achieve very good values above 2.

A t-test for independent samples showed significant differences in attractiveness ( $p = .009$ ), perspicuity ( $p = .044$ ), stimulation ( $p < .005$ ) and novelty ( $p = .007$ ). In these 4 categories the VR app stands out from the screen-based app. There are no significant differences in efficiency ( $p = .444$ ) and dependability ( $p = .256$ ).

### 3.2 Knowledge Acquisition (analysis of error rates)

In all three groups an increase in knowledge (reduction of errors in the created sequence) was measured directly after the training and a decrease with the second measurement repetition one week later.

	pre	difference	post	difference	scnd-post
VR	32.518	57%	13.984	-26%	18.9
SB	29.663	55%	13.236	-30%	18.981
TP	30.984	88%	3.651	-75%	14.555

Tab.1: Mean values of error rates per sample, the lower the number, the better the learning

The strongest increase in knowledge with 88% after training (post) and the strongest decrease in knowledge (scnd post, one week later) with -75% was measured in the TP group.

The TP group has the lowest error rate in the last measurement with 14.6 errors, thus approximately 7-8 wrong steps.

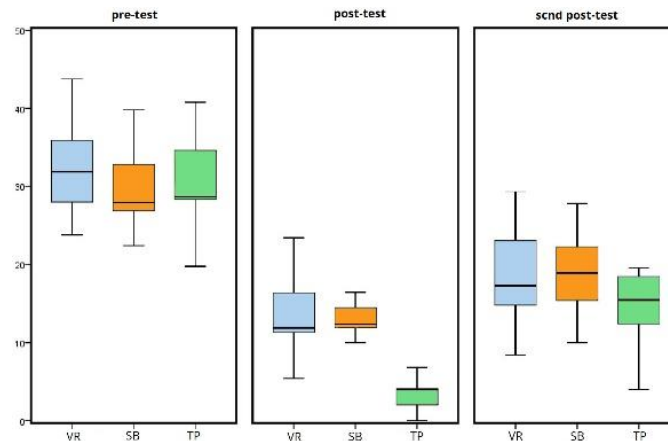


Fig. 3: Error rate, externally separated by measurement time, internally by condition

The error rates were analysed per measurement time using a univariate ANOVA with the between-subject factor *condition* (VR, screen-based, text/picture). The ANOVA of the pre-test revealed no significant differences in the error rates of all samples ( $F(2, 42) = 1,104$ ,  $p = .341$ ). According to Welch-ANOVA of the post-test, there is a statistically significant difference between the groups (Welch test  $F(2, 25,972) = 82,307$ ,  $p < .001$ ).

Games-Howell post-hoc tests show in pair-wise comparison that the error rate of the TP group is significantly lower ( $p < .001$ ) than in the VR and SB group (VR:  $-10.33$ , 95%-CI  $[-14.03, -6.64]$ ; SB:  $-9.59$ , 95%-CI  $[-11.52, -7.65]$ ). There is no significant difference between the two groups VR and SB with an average difference of 0.748 errors. The effect strength  $\eta$  (approximately: significance) according to Cohen [Co88] corresponds with  $\eta = .66$  to a large effect ( $\leq 0.01$  small effect,  $\leq 0.06$  medium effect,  $\leq 0.14$  large effect).

The ANOVA of the second post-test does not indicate a significant influence of the training condition ( $F(2, 42) = 3.104$ ,  $p = .318$ ). However, a slightly lower error rate (VR:  $p = .081$ ; SB:  $p = .074$ ) according to Bonferroni test is noticeable in the text group compared to VR/SB (VR:  $-4.36$ , 95%-CI  $[-9.08, 0.39]$ ; SB:  $-4.43$ , 95%-CI  $[-9.16, 0.31]$ ).

Another measured value was the percentage of correct objects and activities integrated into the built processes in the respective groups at the individual measurement points. Except for one extreme outlier, all participants of the TP group used 100% of the required objects in the post-test. With an average value of 87.18% ( $SD = .129$ ) in VR group and 90.27% ( $SD = .114$ ) in SB group, which corresponds approximately to one forgotten object, the results are just below the TP group.

The mean values in the scnd-post-test have a slightly decreasing tendency. The conditions VR and TP are also in the range of about one forgotten object and SB with one or two missing objects. (VR:  $M = 87.18\%$ ,  $SD = .122$ ,  $n = 15$ ; SB:  $M = 83.19\%$ ,  $SD = .13$ ,  $n = 15$ ; TP:  $M = 92.31\%$ ,  $SD = .122$ ,  $n = 15$ ).

Since no normal distribution can be assumed, Kruskal-Wallis tests are used instead of ANOVA to measure the central trends. The measurement of the post-test shows significant differences both in terms of forgotten objects ( $\chi^2(2) = 8,445$ ,  $p = .015$ ) and activities ( $\chi^2(2) = 10,682$ ,  $p = .005$ ). Paired comparisons with post-hoc tests (Dunn-Bonferroni) reveal that the amount of forgotten activities in both VR ( $z = -3,059$ ,  $p = .007$ ) and SB group ( $z = -2,526$ ,  $p = .035$ ) is significantly higher than in the TP group (effect strength according to Cohen [Co92]: medium effect ( $r = .46$ )). There is no significant difference between VR and SB ( $z = -.533$ ,  $p = .594$ ).

In addition, significantly more objects were forgotten in the VR group compared to the TP group when creating the process steps. The SB group has no significant differences to the other two groups.

In the scnd-post-test, the Kruskal-Wallis test shows no significant differences between forgotten objects ( $\chi^2(2) = 3,987$ ,  $p = .136$ ) and activities ( $\chi^2(2) = 1,793$ ,  $p = .408$ ).

### 3.3 Knowledge Acquisition (analysis of time)

The duration of an infusion preparation in the training simulations averaged 8.3 minutes. This value was rounded up to 9 minutes and set as learning time for the TP condition.

Table 2 shows the average test and training units duration in the respective groups. Although 9 minutes were allowed, the subjects in the TP group completed the learning phase after an average of 6.99 minutes ( $SD = 2.55$ ).

	pre	difference	post	difference	delay	scnd-post	training
VR	7.24	11%	6.46	-27%	7.79	8.91	6.98
SB	7.35	13%	6.42	-7%	8.13	6.92	9.09
TP	8.16	7%	7.55	-8%	7.33	8.21	6.99

Tab. 2: Mean values of time measurements per sample (in minutes), duration post to scnd post test (delay) in days

The time participants required in both training simulations was analysed with a t-test. Accordingly, the infusion preparation was performed significantly faster in VR ( $M = 6.98$ ,  $SD = 1.536$ ,  $n = 15$ ) than in the SB group ( $M = 9.09$ ,  $SD = 1.935$ ,  $n = 15$ ) ( $t(28) = -3.311$ ,  $p = .003$ ). The strength of the effect according to Cohen [Co92] is  $r = .53$  and thus corresponds to a strong effect ( $r = .10$  weak effect,  $r = .30$  medium effect,  $r = .50$  strong effect). As shown in column *delay* in table 2 the actual distance between the first and second post-test in each group exceeds 7 days on average. The greatest distance occurred in the SB group with a maximum of 10 and an average of 8.13 days ( $SD = .99$ ).

The times of the test runs must be examined with regard to their influence on the error rate of the respective test. A Kruskal-Wallis test is used to identify significant differences within the conditions per knowledge test run. Due to the results there are no significant differences in the distribution of the consideration time across all three groups in pre-test ( $\chi^2(1) = .269, p = .910$ ), post-test ( $\chi^2(1) = .590, p = .622$ ) and scnd-post-test ( $\chi^2(1) = 4.056, p = .186$ ). Spearman's rank correlation is used to test whether a low error rate correlates with a long consideration time. In pre- and scnd-post-test, there is a weak tendency for a slightly negative linear correlation according to the rank correlation coefficient (higher time for consideration leads to lower error rate). In pre-test ( $r_s = -.230, p = .129, n = 45$ ) and scnd-post test ( $r_s = -.212, p = .162, n = 45$ ) as well as in post-test ( $r_s = -.109, p = .477, n = 45$ ) the values are not significant.

## 4 Discussion

### 4.1 Usability

The result of the UEQ evaluation corresponds to the expected result when comparing a well-known technology with an innovative and supposedly more exciting one. The point-and-click app has good results on its own, which highlights an adequate implementation of this genre. Nevertheless, it clearly loses in comparison to the VR app, which still has a kind of innovation bonus in terms of user satisfaction and offers a generally more exciting experience.

The dependability of both apps is positively assessed, which supports the results in the following learning effect evaluation. If greater problems had occurred here, higher cognitive load [SVMP98] could have had a significant negative impact on learning outcome. With a few weak exceptions, however, this was not evident in any condition neither through observation during the study nor through statistics.

### 4.2 Learning Effectiveness

The equality of the pre-tests in each group ensures a generally valid baseline in level of knowledge, therefore the measured improvement after the training is well comparable. Since no correlation between the duration of the knowledge test units and the error rate could be found, it can be ruled out that a different test duration leads to a biased result in favor of those who spent more time on the task. Unfortunately not all scnd post-tests were executed after exactly 7 days. Nonetheless an approximately equal distribution in the three groups relativizes this to a large extent. Just the TP group has a slightly earlier time of measurement repetition on average, which does not weaken the results. Assuming that most of the knowledge is forgotten in the first few days, as postulated by Ebbinghaus [Eb85], the difference after 7, 8 or 9 days is negligible.

Considered individually, each group has not only an average but also an absolute lower error rate after the training session. One week later, as expected, some content is forgotten. In comparison, the outstanding difference between the two virtual training simulations and the TP condition in the first post-test is striking. The increase of knowledge is clearly visible in each condition but is significantly lower in the simulation groups. Finally, it is not surprising that there is a clear difference, since the participants of the TP group had 9 minutes to memorize the 23 steps, whereas the participants of the VR and SB conditions could only see each step once. In addition, the TP group had the opportunity to learn all contents until shortly before the time expired, while in the simulation conditions the first steps (due to iterative performance) were already a few minutes ago. In the minimum time span between the end of the training and the post-test, learned contents are hardly forgotten, but time spans around 5-10 minutes can play a decisive role, especially if each step is performed only once [Eb85].

On the other hand, the participants of the simulation groups were able to perform the steps themselves within the scope of the interaction possibilities, thus using a further learning channel and dealing with the respective contents more detailed. Nevertheless, pure memorization is more advantageous in this study, as reflected in the statistics.

A comparison of the VR and the SB group shows no statistically significant difference in general. The small difference in slope (57% / 55%) is far from sufficient to certify a group as "more effective". Similarly, a look at the scnd-post-test shows no significant difference (drop 26% / 30%) between the two simulation groups. The data collected and experience gained during the study hardly allow any room for argumentation contrary to the statistics. Thus, it could not be shown that training in VR facilitates the acquisition of declarative knowledge compared to a related screen-based application. However, the sample size must be taken into account, which was rather small with 15 subjects per group.

Nevertheless, the data of the scnd-post-test shows another anomaly. After one week, the determined knowledge level of the TP group has clearly reduced. Compared to the VR and SB group, TP still has better values, but statistically significant differences are no longer present. This suggests that learning content in training simulations is more efficient in the long run. The comparatively strong knowledge loss of the TP Condition supports this thesis, since the VR and SB groups show just a small loss. Contrary to this assumption, the simulation conditions had less knowledge to lose from the beginning. The TP group had learned a larger number of steps in the training, which is more difficult to remember than the already faulty results of the other groups. In addition, the TP group with evaluation of the scnd-post-test is in the same range of values as VR and SB but has not turned out to be worse. This also allows the thesis that this level of knowledge represents a kind of limit to which test persons fall back after a few days, no matter what kind of training they have completed and how much knowledge they have acquired. The question would thus be whether the users of the training simulations, if they had learned more (e.g. by doing the training twice or three times), would also have forgotten more after one week.

### 4.3 Conclusion

The results leave room for further research, in which the groups "training simulation" and "text/picture" should have the same level of knowledge not only before but also directly after the training session. On the basis of the conducted study, it can be hypothesized that students who have learned all steps of a certain, previously unknown sequence in a training simulation retain the factual steps significantly longer in their memory than persons who have learned it by heart via textual instructions. Another approach could also consider a hybrid solution that takes advantage of the rapid acquisition of knowledge through text/picture instructions and then consolidates what has been learned through a training simulation. According to the results of this study, a differentiation of the simulation types between VR and Screen-Based (related to the acquisition of declarative knowledge) does not seem to be necessary anymore, but a repeated differentiation might, although not expected, provide a different result.

### Bibliography

- [Am01] Amory, Alan: Building an educational adventure game: Theory, design, and lessons. *Journal of Interactive Learning Research*, 12(2):249–263, 2001.
- [BMS88] Barbe, Walter Burke; Milone, Michael N; Swassing, Raymond H: Teaching through modality strengths: Concepts and practices. Zaner-Bloser, 1988.
- [Co88] Cohen, Jacob: Statistical power analysis for the behavioral sciences. Technical report, 1988.
- [Co92] Cohen, Jacob: A power primer. *Psychological bulletin*, 112(1):155, 1992.
- [Eb85] Ebbinghaus, Hermann: Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Duncker & Humblot, 1885.
- [LHS08] Laugwitz, Bettina; Held, Theo; Schrepp, Martin: Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and Usability Engineering Group. Springer, pp. 63–76, 2008.
- [MVM09] Miller, Frederic P.; Vandome, Agnes F.; McBrewster, John: Levenshtein Distance: Information Theory, Computer Science, String (Computer Science), String Metric, Damerau-Levenshtein Distance, Spell Checker, Hamming Distance. Alpha Press, 2009.
- [Pf18] Pfeiffer, T. et al.: Virtual SkillsLab-Trainingsanwendung zur Infusionsvorbereitung (Wettbewerbssieger). In: DeLFI Workshops 2018. Proceedings der Pre-Conference-Workshops der 16. E-Learning Fachtagung Informatik (DeLFI 2018), Frankfurt, Germany, September 10, 2018. volume 2250, 2018.
- [SVMP98] Sweller, John; Van Merriënboer, Jeroen JG; Paas, Fred GWC: Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296, 1998.
- [Xi20] Xin, B. et al.: The efficacy of immersive virtual reality surgical simulator training for pedicle screw placement: a randomized double-blind controlled trial. *International Orthopaedics*, pp. 1–8, 2020.

## Pre-Testing der Lerneffektivität von 2D-to-3D-Didactics in immersiven VR-Umgebungen

Carsten Fussan<sup>1</sup>, Veronika Hanesová<sup>2</sup>

**Abstract:** Sind Lehrdarstellungen in 3D-Virtual Reality (VR) klassischen zweidimensionalen Lehrmedien in universitären betriebswirtschaftlichen Studiengängen überlegen? In Vorbereitung einer empirischen Studie zu diesem Thema wurde ein Pretest durchgeführt. Hierbei wurde festgestellt, dass ein empirischer Vergleichstest zur Prüfung der Überlegenheit von VR nur valide erfolgen kann, wenn sowohl Vorerfahrungen der Probanden mit VR-Medien, als auch ihre jeweilige individuelle Adaptionsfähigkeit an VR berücksichtigt werden. Der Artikel gibt einen Überblick über die Konzeption, die Technik, die Durchführung sowie die Ergebnisse des Pretests.

**Keywords:** Lerneffektivität, 2D-to-3D-Didactics, 3D-to-3D-Didactics, Virtual Reality, Virtual Reality Lernen, VR, VR Lernen, VR Lehre, Immersion

### 1 Einleitung

Die Übertragung von 2D-darstellbaren Lehrinhalten in 3D-VR<sup>3</sup>-Anwendungen erscheint besonders interessant: Wenn die Anwendung von 3D-Virtual Reality (VR) in der Lehre eine Breitenakzeptanz erfahren soll, dann muss diese Technologie hinsichtlich der Lernwirkung vor allem traditionellen 2D-Lehrdarstellungen überlegen sein. Die Studiengänge mit den meisten Studierenden in Deutschland sind Betriebswirtschaftslehre (BWL), Informatik und Rechtswissenschaften - zusammen mit einem Studierendenanteil von 30% [SB19]. Wenn in diesen Studiengängen eine Visualisierung von Lehrinformationen erfolgt, so wird diese bisher in 2D vorgenommen (insbesondere: Tafelbilder, Beamer-Projektionen und 2D Online-Systeme). Wenn VR also ein überlegenes Lehr- und Lernmedium ist, so müsste es auch jenen 2D-Darstellungen in den genannten Studiengängen überlegen sein. In den bisherigen empirischen Studien lässt sich eine Tendenz erkennen, welche diese These bestätigt, ein repräsentativer Beweis hierfür steht jedoch noch aus (vgl. Abschnitt 2.2). Zudem wäre es auch nicht ausreichend zu wissen, dass VR (möglicherweise) der klassischen 2D-Lehre überlegen ist, es ist insbesondere wichtig zu wissen, welche didaktischen Elemente innerhalb der VR- diese Überlegenheit auslösen. Nur dann können Anwendungen optimiert und wirkungsgerecht programmiert

---

<sup>1</sup> Hochschule Anhalt, Bernburger Straße 55, 06366 Köthen, carsten.fussan@hs-anhalt.de.

<sup>2</sup> Hochschule Anhalt, Strenzfelder Alle 28, 06406 Bernburg, veronika.hanesova@hs-anhalt.de.

<sup>3</sup> Computeranimierte Welt in Echtzeit, die mithilfe spezieller Soft- und Hardware (insb. VR-Brillen) erzeugt wird, wobei die Außenwelt ausgeblendet und menschliche Sinne durch Videos, Animationen und Audio beeinflusst werden [Do13].

werden. Die nachfolgenden Ausführungen beschreiben Erkenntnisse aus einer Pretest-Untersuchung zur Vorbereitung einer repräsentativen empirischen Studie zur Ermittlung der Lerneffektivität von VR im Lehrgebiet der BWL.

## 2 Forschungskontext

### 2.1 Definitionen

Die Steigerung der *Lerneffektivität* bezeichnet eine Verbesserung des Lernergebnisses bei ansonsten unverändertem inhaltlichem Umfang der gelehnten wissensrelevanten Informationen. Insbesondere bezüglich der Überlegenheit von VR-Anwendungen gegenüber traditionellen Formen des Lernens [MN11], [Ze18] geht es im analytischen Kern um die Frage, ob durch diese innovativen Techniken eine Steigerung der Lerneffektivität ausgelöst wird.

Als *3D-to-3D-Didactics* werden nachfolgend VR-Lehranwendungen bezeichnet, welche in der Realität originär dreidimensionale Umgebungen/Vorlagen in eine ebenso dreidimensionale VR-Lehrdarstellung transportieren. Beispiele hierfür sind die 3D-Simulation gefährlicher Arbeitswelten oder die 3D-Darstellung unzugänglicher Umwelten zum Zwecke wirklichkeitsnaher Erlebnisse.

Als *2D-to-3D-Didactics* werden nachfolgend VR-Lehranwendungen bezeichnet, welche bisher zweidimensional dargebotene Inhalte in eine dreidimensionale VR-Lehrdarstellung transportieren. Ein Beispiel hierfür ist die 3D-Visualisierung von 2D-Grafiken, Präsentationen und 2D-Bildern aus betriebswirtschaftlichen Lehrmedien.

### 2.2 Empirische Evidenz

Die Wirkung von VR auf die Lerneffektivität wurde bisher aus verschiedenen Perspektiven empirisch untersucht. In einigen Studien wurden Vorteile der virtuellen Lernanwendung gegenüber realen Vermittlungsformen identifiziert [We15], [CB19]. In anderen Studien konnten hingegen keine oder negative Effekte von VR auf die Lerneffektivität identifiziert werden [PDN08], [Ha18]. Die meisten empirischen Studien - welche eine positive Lernwirkung von VR feststellten - haben 3D-to-3D-Didactics untersucht [Ba08], [Mo09], [Ma17], [CB19], [K119], [Wi19]. Nur zwei der empirischen Studien hatten 2D-to-3D-Didactics zum Inhalt, in beiden Untersuchungen mit einer positiven Wirkung auf die Lerneffektivität [We15], [DSL19].



### 3 Empirischer Pretest im Bereich 2D-to-3D-Didactics

#### 3.1 Technisches und didaktisches Setup der 3D VR Testumgebung

Zur Vorbereitung der empirischen Hauptuntersuchung wurde ein Pretest mit einem hierfür programmierten VR-Prototyp durchgeführt. Diese Testumgebung wurde mittels Unity3D für eine Betrachtung mit der Oculus Rift S erstellt und BWL-Lehrinhalte zur Methodik der „Szenarioplanung“ wurden darin positioniert (2D-to-3D). Mit dem Ziel der bestmöglichen Erreichung von Immersion wurden neben einer wirklichkeitsnahen 3D-Darstellungsqualität auch Animationen sowie ein komplexes Shader-Verhalten für das Auflösen von dreidimensionalen Objekten zum Wechsel von visuellen Raumeindrücken eingesetzt. Unterstützend wurde eine atmosphärisch glaubhafte Klangkulisse erschaffen (immersiver 360°-Sound).



Abb. 1: Hologramm Professor & 3D-Grafik mit Keywords (links), virtueller Avatar (rechts)

In die VR-Testumgebung wurden folgende 7 didaktische Elemente implementiert: Individuelle Nutzerbewegung (Element 1) mit dem Ziel bestmöglicher Rezeption des sukzessiven Aufbaus und der Aussagen der 3D-Graphiken (Element 2), zeitlich abgestimmte Einblendungen von Keywords (Element 3), ein virtueller Avatar zur Begleitung des Lernprozesses (Element 4), ein mittels Avatar projiziertes Image eines Professors mit gesprochenen Erläuterungen der Lerninhalte (Element 5), ein auf die Lerneinblendungen abgestimmter und entsprechend räumlich ausgerichteter Sound (Element 6) und ein virtuelles Notizsystem, in welchem die Probanden mit Hilfe eines Controllers freihändige Notizen erfassen können (Element 7).

#### 3.2 Empirische Ergebnisse des Pretests

Im Dezember 2019 bis März 2020 wurden aus einer im Hochschulumfeld zugänglichen Zielgruppe von Studierenden und berufstätigen Absolventen der BWL zufällig 10 Personen ausgewählt. Die unterschiedlichen betriebswirtschaftlichen und lernrelevanten Leistungsniveaus der Probanden wurden durch Fragen zu Vornoten aus Bachelor-Studiengang (für die Master-Probanden), zur Note der Hochschulzugangsberechtigung (für die Bachelor-Probanden) und zur Abschlussnote des Studiums (für die Absolventen) erfasst. Dies wurde durch eine Frage zur aktuellen Durchschnittsnote des jeweils aktiven Studiums ergänzt. Für die Gruppe der berufstätigen Absolventen wurde das Leistungsniveau

durch Verrechnung aller angegebenen Abschlussnoten beendeter Studiengänge ermittelt. Aus diesen gesamten Angaben wurde für jeden Probanden ein „Pre-Qualification-Index“ (PQI) gebildet, welcher die Einordnung in verschiedene Leistungsgruppen ermöglicht. Alle Probanden sollten zudem angeben, ob sie Vorerfahrungen mit dem Medium VR haben. Zur Erfassung des Lernergebnisses wurden den Probanden insgesamt 7 Kontrollfragen<sup>4</sup> zu den während des Tests vermittelten Lehrinhalten vorgelegt, welche unterschiedliche Verständnisniveaus prüften: einfache Wiederholungsfragen gelernten Wissens bis hin zu Transferfragen intensiven Verständnisses. Die 10 Probanden wurden für die Lehrprobe in 2 zufällige Gruppen eingeteilt. Die Kontrollgruppe erhielt das BWL-Wissen in klassischer Vorlesungsart in einem Hörsaal durch einen Dozenten mit Beamer unter Nutzung derselben Tonspur dargeboten, die auch die Testgruppe innerhalb der VR-Anwendung zu hören bekam. Die Testgruppe wurde mit einer VR-Brille ausgestattet und nur der virtuellen Lehranwendung ausgesetzt. Beide Lehrproben dauerten insgesamt ca. 10 Minuten.

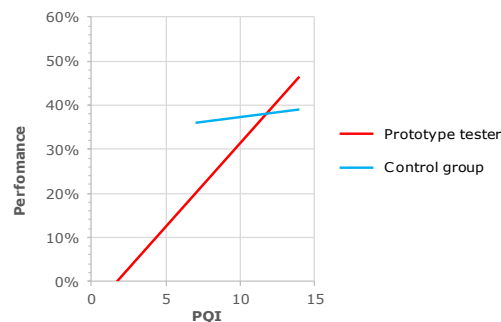


Abb. 2: Regressionsanalyse Auswertung Pre-Test

Nach Regression der Einzelergebnisse ergab sich folgendes Testergebnis: Die Auswertung zeigt eine höhere Lerneffektivität von 2D-to-3D Didactics in der Testgruppe der hoch leistungsfähigen Probanden. In Abbildung 2 ist die Regressionsgerade durch die Einzelergebnisse der VR-Testgruppe rot gefärbt, die der Kontrollgruppe blau. Es ist insbesondere auffällig, dass die Testgruppe die Kontrollfragen umso fehlerhafter beantwortet hatte, je leistungsschwächer die jeweiligen Probanden eingruppiert waren (vgl. PQI-Index). Eine Erklärung für diesen Kurvenverlauf lieferten die Aufzeichnungen der Versuchsleiter, die parallel zu den einzelnen Tests angefertigt wurden und die danach jeweils in Verbindung mit den Angaben der Teilnehmer zu ihren Vorerfahrungen mit VR gesetzt wurden.

Je geringer die Vorerfahrungen der Probanden mit VR waren, desto schwieriger war es für die Testpersonen, sich in die unbekannte VR-Lehrumgebung hineinzufinden und sich nach einer kurzen Betrachtung des virtuellen Raumes unbeeinflusst von den neuen und faszinierenden 3D-Stimuli auf die dargestellten Lehrinhalte zu konzentrieren. Die Testpersonen beklagten dies in umgekehrt proportionaler Intensität zu ihrer Leistungsperformance nach Abschluss und auch während des Tests. Einige VR-unerfahrene Probanden

<sup>4</sup> Auf Basis der Taxonomie von Bloom [B154].

baten um Wiederholung der VR-Vorlesung. Verstärkt wurde dieser Effekt differenter Vorerfahrungen durch sehr unterschiedlich ausgeprägte individuelle Fähigkeiten der Probanden, sich an die unbekanntere Lehrumgebung zu adaptieren. Einige als hoch leistungsfähig eingruppierte Probanden brauchten deutlich weniger Zeit zur Adaption an die VR-Testumgebung als Testpersonen mit einem geringeren PQI – bei jeweils vergleichbar langen 3D-Vorerfahrungen.

Insgesamt ist festzustellen, dass eine valide Testung der Lerneffektivität von 2D-to-3D-Didactics aufgrund des störenden Einflusses der individuellen VR-Vorerfahrungen der Probanden und ihrer unterschiedlichen individuellen Adaptionen an VR-Umgebungen mittels der für den Pretest genutzten VR-Anwendung nicht möglich ist.

#### **4 Fortgang der empirischen Untersuchungen**

Die Vorstudie hat keinen repräsentativen Charakter, gibt jedoch wesentliche Hinweise für die Durchführung der Hauptstudie. Ein wichtiges abgeleitetes Ziel für die Durchführung der Hauptuntersuchung ist die Schaffung vergleichbarer Ausgangspositionen der Konzentration auf die Lehrinhalte bei allen Testpersonen.

Folgende Punkte werden an der Teststruktur für die Hauptuntersuchung geändert:

- Die Vorerfahrungen der Probanden mit VR-Medien werden deutlich stärker beachtet. Diese sollen durch mehrere Indikatoren erfasst und vom PQI getrennt ausgewertet werden. Ziel ist es, nur Probandengruppen gleicher VR-Vorerfahrungen hinsichtlich der Lerneffektivität miteinander zu vergleichen.
- Die unterschiedlichen individuellen Adaptionen der Probanden an die neue Lehrumgebung sollen ausgeglichen werden. Hierfür erhält die VR-Testumgebung zwei technische Erweiterungen: die Phase der Gewöhnung im 3D-Raum soll deutlich verlängert und vereinfacht werden und die Testpersonen sollen selbst entscheiden können, wann sie bereit sind, sich voll auf die Lerninhalte zu konzentrieren, indem sie den eigentlichen VR-Vorlesungsteil selbst starten.
- Der Stichprobenumfang soll auf 120 bis 150 Probanden ausgedehnt werden, so dass Repräsentativität nachgewiesen werden kann.
- Es sollen 7 didaktische VR-Elemente hinsichtlich ihrer Einzelwirkung auf die Lerneffektivität untersucht werden.

#### **Literaturverzeichnis**

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

- [Bl54] Bloom, B. S.: Taxonomy of educational objectives: The Classification of Educational Goals, 1954.
- [CB19] Checa, D.; Bustillo, A.: Advantages and limits of virtual reality in learning processes: Brivesca in the fifteenth century, In: Virtual Reality, Band 24, S. 151-161, 2019.
- [Do13] Dörner, R. et.al.: Virtual und Augmented Reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität, 2013.
- [DSL19] Demitriadou, E.; Stavroulia, K.; Lanitis, A.: Comparative evaluation of virtual and augmented reality for teaching mathematics in primary education, In: Education and Information Technologies, Band 25, S. 381-401, 2019.
- [Ha18] Harrington, C. et.al.: 360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention, In: Journal of Surgical Education, Band 75, Nr. 18, S. 993-1000, 2018.
- [Kl19] Klingauf, A. et.al: Wirkung von interaktiven 3D-360°-Lernvideos in der praktischen Ausbildung von Handwerkern, In: Die 17. Fachtagung Bildungstechnologien, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik Bonn, S. 145-156, 2019.
- [Ma17] Madathil, K. et.al.: An Empirical Study Investigating the Effectiveness of Integration Virtual Reality-based Case Studies into an Online Asynchronous Learning Environment, In: Computers in Education Journal, Band 8, Nr. 3, S. 1-10, 2017.
- [MN11] Mikropoulos, T.; Natsis, A.: Educational virtual environments: A ten-year review of empirical research (1999–2009), In: Computers & Education, Nr. 56, S. 769-780, 2011.
- [Mo09] Mount, N. et.al.: Learner immersion engagement in the 3D virtual world: principles emerging from the DELVE project, In: Innovation in Teaching and Learning in Information and Computer Sciences, Band 8, Nr. 3, S. 40-55, 2009.
- [PDN08] Patera, M.; Draper, S.; Naef, M.: Exploring magic cottage: a virtual reality environment for stimulating children’s imaginative writing, In: Interactive Learning Environments, Band. 16, Nr. 3, S. 245-263, 2008.
- [SB19] Statistisches Bundesamt: Anzahl der Studierenden an deutschen Hochschulen in den 20 am stärksten besetzten Studienfächern im Wintersemester 2018/19, <https://de.statista.com/statistik/daten/studie/2140/umfrage/anzahl-der-deutschen-studenten-nach-studienfach/>, [abgerufen 24. März 2020].
- [Wi19] Wiepke, A. et.al.: Einsatz von Virtual Reality zum Aufbau von Klassenmanagement-Kompetenzen im Lehramtsstudium, In: Die 17. Fachtagung Bildungstechnologien, Lecture Notes in Informatics (LNI), Gesellschaft für Informatik Bonn, S. 133-144, 2019.
- [We15] Webster, R.: Declarative knowledge acquisition in immersive virtual learning environments, In: Interactive Learning Environments, Band 24, Nr. 6, S. 1319-1333, 2015.
- [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), 2018.

## Demokratische Schule 360°: Potenziale Virtueller Realität für LehrerInnenbildung und Ethnografie

Tobias Leßner<sup>1</sup>

**Abstract:** Das Aufkommen von immer günstigeren VR-Brillen und 360°-Kameras eröffnet neue Einsatzmöglichkeiten in Lehrerbildung<sup>2</sup> und ethnografischer Forschung. Im laufenden Projekt *Demokratische Schule 360°* bilde ich den Alltag einer Demokratischen Schule als Virtuelle Realität (VR) ab und setze diese in meinen Grundschulpädagogik-Seminaren ein. Studierende sollen so die Möglichkeit erhalten, immersive und ethnografische Einblicke in den Alltag der abgebildeten Schule zu gewinnen. Im vorliegenden Artikel verknüpfe ich in sechs Abschnitten jeweils zwei der vier Themen des Projektes (VR, Demokratische Schule, Lehrerbildung, Ethnografie), sodass am Ende ein differenziertes Gesamtbild des Projektes im Spannungsfeld von Vision und Wirklichkeit entsteht. Dabei werden vor allem Fragen aufgeworfen, statt Antworten gesucht.

**Keywords:** Virtuelle Realität, 360°-Aufnahmen, Lehrerbildung, Ethnografie, Demokratische Schule

### 1 Einleitung

Die ursprüngliche Idee für dieses erziehungswissenschaftliche Projekt entstand 2017 im Rahmen eines universitätsweiten Förderwettbewerbes zur Verbesserung von Lehre und Studium und wird entsprechend aus Qualitätsverbesserungsmitteln gefördert. Ziel ist es, den Lehramtsstudierenden das theoretische Konzept und den pädagogischen Alltag Demokratischer Schulen näher zu bringen. Zum Erreichen dieses Ziels war zunächst vorgesehen, einen Pool an videoethnografischem Datenmaterial (noch ohne 360°-Videos und ohne VR) zu erstellen, damit die Studierenden z.B. eine Schulversammlung in Gänze anschauen oder das Team der Schulküche beim Kochen und Schüler beim Lernen beobachten können<sup>3</sup>. Mithilfe der Videos sollten authentische Einblicke in den Alltag der Demokratischen Schulen ermöglicht werden – ähnlich den Einblicken, die man während einer Hospitation gewinnt. In der Weiterentwicklung des Projektes und mit dem Aufkommen günstiger VR-Technik<sup>4</sup> 2018 änderte sich zwar diese grundlegende Ausrichtung nicht, jedoch der Weg und das Ziel der Umsetzung: Im Fokus steht nunmehr die Darstellung der beforschten Demokratischen Schule als VR, um ein immersives Erleben des Schulalltags zu ermöglichen. Die Auseinandersetzung mit Ähnlichkeiten und Differenzen von VR als

---

<sup>1</sup> Universität Siegen, AG Grundschulpädagogik, Adolf-Reichwein-Str. 2, 57068 Siegen, tobias.lessner@uni-siegen.de

<sup>2</sup> Zur besseren Lesbarkeit verwende ich das generische Maskulinum, meine damit aber alle Geschlechter.

<sup>3</sup> Bereits existierende Videos, Filme und Erfahrungsberichte [bspw. HT15] zeigen weniger den Schulalltag in vivo als vielmehr, wie die Akteure ihren Alltag vorrangig in Interviews präsentieren.

<sup>4</sup> Im Projekt finden Verwendung: Oculus Go, Ricoh Theta V, Pano2VR sowie VR Tourviewer.

Bildungsvision und dem Schul- bzw. Seminaralltag ist dabei konstitutiv für dieses Projekt. Es geht um eine Annäherung an eine Utopie von VR.

An dieser Stelle unabhängig von verschiedenen Definitionen meint VR in diesem Projekt eine Verknüpfung von 360°-Fotos und -Videos der schulischen Räume zu einer virtuellen Umgebung. Diese virtuelle Umgebung wird vom Rezipienten mittels einfacher Pointer-Klick-Steuerung in einer VR-Brille der Raumlogik der Schule folgend *durchlaufen*. Statt jedoch physisch durch eine Tür zu gehen, klickt der Rezipient diese an und gelangt so zur Aufnahme des nächsten Raumes. Die 360°-Aufnahmen der jeweiligen Räume sind dann ergänzt durch einzeln aufrufbare Infotexte, Fotos, Videos, ethnografische Feldnotizen und Audio-Elemente. Der Raum wird lebendig, sodass der Rezipient an den sozialen Situationen teilhaben kann. Was genau *Teilhabe* in diesem Kontext bedeutet und wie sich diese von einer Hospitation unterscheidet, wird noch zu beschreiben sein.

Das Ziel des hier vorliegenden Artikels ist es, grundlegende Fragen, die sich im Laufe des Projektes bisher ergeben haben, systematisch zu illustrieren, um so das Potenzial des Projektes in den sechs (nicht trennscharfen) Verknüpfungen der vier Themen aufzeigen zu können. Dabei ist die Datenerhebung abgeschlossen und eine erste Version der virtuellen Umgebung erstellt, wobei diese fortlaufend den Erkenntnissen aus der andauernden Analyse angepasst wird. Die virtuelle Umgebung konnte aufgrund der aktuellen Verpflichtung zur digitalen Lehre aus datenschutzrechtlichen Gründen lediglich mit einzelnen Akteuren erprobt und noch nicht wie geplant im Seminarkontext eingesetzt werden.

## 2 Lehrerbildung und Demokratische Schule

Ausgangspunkt der initialen Projektidee war die Suche nach einem hochschuldidaktischen Weg, „um die aus Erfahrung resultierenden Wissensstrukturen und Einstellungen [der Lehramtsstudierenden in Bezug auf Schule und Unterricht durch ihre eigene Schülerbiografie, TL] zu **dekonstruieren**’ und dafür zu sorgen, dass die Lernenden ihre Wissensstrukturen und Einstellungen aktiv und neu konstruieren, so dass diese tragfähig und handlungsleitend für ihre Berufsbiographie werden“ [MKH07]. Die pädagogischen Konzepte Freier Demokratischer Schulen von Schule und Unterricht<sup>5</sup> stehen in starkem Kontrast zur Pädagogik der meisten staatlichen Schulen. Das Bild der Lehramtsstudierenden von Schule und Unterricht, die meist nicht an diesen Demokratischen Schulen sozialisiert sind, kann *so auf den Kopf gestellt* werden. Es soll jedoch nicht darum gehen, in einer normativ aufgeladenen Art und Weise diese pädagogischen Ideen Demokratischer Schulen zu propagieren, sondern über das Aufspüren von Differenzen in ein konstruktiv-produktives Gespräch über Schule und Unterricht zu kommen. Aus einer „Befremdung der eigenen Kultur“ [AH97], bzw. hier: der eigenen *Schulkultur*, wird im Projekt eine Kontrastierung der

---

<sup>5</sup> Demokratische Schulen sind Schulen in freier Trägerschaft. In Deutschland existieren je nach Definition ca. 15 dieser Schulen. Kern des pädagogischen Konzeptes ist ein sehr hohes Maß an Beteiligung der Lernenden an den Fragen der Schulgestaltung und den Fragen des eigenen Lernprozesses. Ausführlicher [K113].

eigenen mit einer den Studierenden fremden Schulkultur. Bei diesem rekurrierenden Vergleich, beim Eintauchen in den Alltag Demokratischer Schulen und bei der Auseinandersetzung mit dieser Schulkulturalternative entsteht die Frage: Wie greifen Lehramtsstudierende die pädagogischen Ideen Demokratischer Schulen auf, bearbeiten diese und fügen sie dann in ihre bestehenden Wissensstrukturen und Einstellungen ein?

### 3 Lehrerbildung und VR/360°

Anders als beispielsweise bei Johnson [Jo18] schauen die Studierenden keine einzelnen, kurzen, frei verfügbaren 360°-Videos, sondern betreten die eigens für diesen Zweck konstruierte virtuelle Umgebung. Im Hinblick auf die Notwendigkeit der Digitalisierung von Schule und Lehrerbildung können sie das Medium VR nicht nur als Rezipienten erproben, sondern erhalten zudem die Möglichkeit, die VR mit der entsprechend zur Verfügung gestellten Software zu gestalten. Die virtuelle Umgebung ist somit auch ein Lernmedium ihrer digitalen Professionalisierung. Wie kann ein Übertrag der im Seminar erlernten digitalen Aspekte auf den späteren Berufsalltag als Lehrer erfolgen?

Beim Einsatz der virtuellen Umgebung, welcher bisher nur mit einzelnen Studierenden und Kolleginnen stattfinden konnte, stehe ich vor vielfältigen pragmatischen Herausforderungen des Zeigens. Mangelnde Verfügbarkeit von Geräten, noch nicht vorhandene Bedienkompetenzen oder Probleme wie Übelkeit und Kopfschmerzen erfordern flexible und situative didaktische Lösungen. Auch erzeugen VR-Brillen eine neue Art des sozialen Miteinanders, wenn die physisch im Raum Anwesenden nicht mehr visuell (und teils wegen Kopfhörern auch nicht mehr auditiv) interagieren können. Die Rezipienten stehen vor der Herausforderung, ihre gleichzeitige An- und Abwesenheit im Seminarraum zu bearbeiten. Aus einer medienethnographischen Perspektive gefragt: Wie bringen sie die Realität mit der virtuellen Realität in sozialen Praktiken miteinander in Einklang? Erste Beobachtungen lassen verschiedene Praktiken des *Sich-Verabschiedens* („Bis gleich“) und des *Sich-Begrüßens* („Da bin ich wieder“) vermuten. Dieser Ein- und Ausstieg in die virtuelle Umgebung wird dabei vor allem von Anfängern stärker inszeniert. Zudem müssen die Rezipienten mit verschiedenen Erzähl- und Berichtspraktiken einander erklären, was sie just virtuell erleben.

Letztlich muss sich die virtuelle Umgebung an den Alternativen messen lassen: Wie unterscheidet sie sich (aus Sicht des Rezipienten) vom Medium Film, Video oder Text? Wie beeinflusst das Medium VR die Auseinandersetzung mit den Inhalten? Wie wirken sich dabei äußere Einflüsse auf das immersive Erleben aus?

### 4 Lehrerbildung und Ethnografie

Im Zentrum ethnografischen Forschens steht der Alltag der Menschen: „Es ist das Alltägliche mit seinen innewohnenden Regeln, seinen Ordnungen und Ritualen. Die Menschen

bringen ihren Alltag gemeinsam hervor. Das wissenschaftlich begründete Verstehen-Lernen dieses Alltags der Menschen, seien es Kinder oder Erwachsene, Lehrer oder Schüler, Polizisten oder Gesetzesbrecher, Wissenschaftler oder Medizinmänner, Rapper oder Banker, ist das Ziel einer ethnographisch angelegten Sozialwissenschaft“ [Wi11]. Auch eine ethnografische Forschungshaltung kann somit zu einer Dekonstruktion des Bildes von Schule beitragen. Wiesemann [Wi11] beschreibt dies weiter wie folgt: „Aus Sicht der Lehrerinnen und Lehrer besteht der didaktische Gewinn dieser Forschungen vor allem in der Übung des ´ethnographischen Blicks´. Es steht also erleichternder Weise nicht die Frage im Mittelpunkt: Mache ich guten Unterricht? sondern: Welcher Unterricht passiert hier? Welche Lernkultur hat sich hier etablieren können und welchen Beitrag leisten die Schülerinnen und Schüler dazu, und welchen leisten die Lehrerinnen und Lehrer? Welche institutionellen und materiellen Zwänge werden sichtbar? Von Handlungszwängen weitgehend entlastet, bieten ethnographische Analysen von real existierendem Unterricht die Möglichkeit, alltägliche Praxis in konkreten Klassenzimmern zu überdenken und möglicherweise konkrete Lernarrangements zu verändern.“

Es wird noch zu erproben und zu evaluieren sein, inwieweit die Studierenden als Rezipienten der VR diese auch als (angeleitete) ethnografische Beobachter betreten und selbstständig analysieren werden. Alternativ kann die VR auch Forschungsergebnisse präsentieren und den Rezipienten durch die bereits getätigte, ethnomethodologische Analyse des Schulalltags führen. Nur, wird die virtuelle Umgebung dann nicht zu einer Art DVD-Menü, wie es bei kamera-ethnografischen Forschungen üblich ist?

## 5 Demokratische Schulen und Ethnografie

Die Ethnografie als Forschungsstrategie, mit der das *Wie* alltäglicher sozialer Praxis beobachtbar wird, eröffnet vielfältige Optionen der „Befremdung des scheinbar Vertrauten“ [AH97] und ermöglicht einen neuen Blick auf die scheinbar selbstverständlichen Phänomene des Alltags an der beforschten Schule [vgl. Le20]. Dieses *Wie* steht dabei im Zentrum meines Promotionsprojektes, an welches das hier beschriebene Projekt gekoppelt ist und welches ich an einer weiteren Demokratischen Schule durchführe. Einen Forschungsstand zu den Praktiken an Demokratischen Schulen gibt es nicht, sodass sowohl das hier beschriebene Projekt *Demokratische Schule 360°* als auch mein Promotionsprojekt diese Desiderata der Frage nach der Hervorbringung des Alltags erstmals bearbeiten.

## 6 Demokratische Schulen und VR/360°

Wenn es aus unterschiedlichen Gründen nur schwerlich möglich ist, die Studierenden an die Schulen zu bringen, dann muss die Schule zu den Studierenden gebracht werden – so der Gedanke zu Beginn des Projektes. Beim Filmen standen wir dann vor einer Reihe ganz pragmatischer Probleme. Täglich mussten wir z.B. überlegen, wie wir in einer Schule, in



der es keine Klassenstrukturen gibt und in der sich jeder jederzeit frei bewegen kann, Situationen einfangen, nicht jedoch aber diejenigen, die nicht gefilmt werden wollten oder durften. Zumal es bei einer 360°-Kamera kein Davor und Dahinter gibt, also keinen Raum, in dem sich die Akteure bewegen können, ohne gefilmt zu werden. Sie können sich nicht hinter der Kamera verstecken. Wie, mit welchen Ethno-Methoden [Ga84], bringen Schüler, Mitarbeiter und die Ethnographen gemeinsam *360°-Grad-Filmen* hervor? Forschungsethische Aspekte und Fragen des Datenschutzes schließen sich an. Wie, bezogen auf das Endprodukt *virtuelle Umgebung*, können wir den Rezipienten dabei von einem heimlichen Zuschauer zu einem Teilnehmer an diesen Situationen machen? Wie kann das Gefühl der Immersion perfektioniert werden? Es geht dabei nicht etwa um das Erleben von bspw. Höhenangst, indem der Rezipient in der virtuellen Umgebung an einer Klippe steht. Vielmehr geht es um das Involviertsein in die sozialen Praktiken der Schule. Welche konkreten Themen und Gegenstände reformschulischer Praxis eignen sich für die Bearbeitung mit VR und 360°-Aufnahmen und wie kann die Bearbeitung evaluiert werden?

## 7 VR/360° und Ethnografie

Wie passen VR und 360°-Aufnahmen zu einer Ethnografie, die sich bereits verschiedener verfügbarer Medien der Dokumentation, Analyse und Darstellung bedient [vgl. Br15]? Ein an einem Thema oder Gegenstand, z.B. an Schulregeln [Le20], festgemachter Vergleich zu bisher eingesetzten Medien ist notwendig. An dieser Stelle lässt sich beispielhaft fragen: Welche methodischen Fragen werden beim Einsatz einer 360°-Kamera aufgeworfen, wie etwa die Frage nach der Rolle des Ethnografen? Bei einer 360°-Kamera ohne ein Davor und Dahinter ist nun auch der Ethnograf im Bild und nimmt vermeintlich anders an der Situation teil als mit einer handgeführten Kamera. Da eine 360°-Kamera auch nicht bewegt werden sollte, da dies später beim Betrachten in einer VR-Brille schnell zu Übelkeit und Schwindel führen kann, steht diese still auf einem Stativ: Was macht der Ethnograf?

Zudem müssen die neuen Technologien methodisch und methodologisch verortet werden. Welche Überlegungen der Videografie, der Kameraethnografie, ethnomethodologischen Videografie, Visual Ethnography, Digital Ethnography oder der Virtual Ethnography sind für das Projekt relevant?

## 8 Ausblick

Es soll im Projekt *Demokratische Schule 360°* zunächst darum gehen, weitere didaktische, methodische und methodologische Fragen zu allen Verknüpfungen der Themen aufzuwerfen und diese detailliert zu formulieren. In Seminaren und Workshops, in frei zugänglichen Ausstellungen und Online-Darstellungen sowie beim Erstellen von virtuellen Umgebungen anderer Felder sollen diese Fragen dann nach und nach beantwortet werden.

Der Fortgang des Projektes ist auch von den technologischen Fortschritten abhängig: höhere Videoauflösungen beim Filmen, bessere Displays, größere Sichtfelder, sowie neue Softwarefunktionen beim Bearbeiten und beim Erstellen von VR-Inhalten werden vermutlich direkte Verbesserungen der Immersion zur Folge haben. Wie können zudem weitere Technologien wie Chatbots oder Animationen die virtuelle Umgebung bereichern? Wenn die Zukunft den Wearables gehört, können wir dann zukünftig Demokratische Schulen und Lehramtsstudierende zu virtuellen Echtzeithospitationen als Hologramme in Augmented- oder Mixed-Reality zusammenbringen? Mit z.B. der *Microsoft HoloLens 2* sind jedenfalls schon heute Produkte auf dem Markt, die dies ermöglichen. Die Potenziale von VR und 360°-Aufnahmen sind unübersehbar. Sowohl aus der Perspektive der Lehrerbildung als auch der Ethnographie bieten sich viele Möglichkeiten, alte Fragen neu aufzurollen und neue zu stellen.

## Literaturverzeichnis

- [AH97] Amann, K.; Hirschauer, S.: Die Befremdung der eigenen Kultur. Ein Programm. In (Hirschauer, S.; Amann, K. Hrsg.): Die Befremdung der eigenen Kultur. Zur ethnographischen Herausforderung soziologischer Empirie. Suhrkamp, Frankfurt am Main, 1997; S. 7–52.
- [Br15] Breidenstein, G. et al.: Ethnografie. Die Praxis der Feldforschung. UVK Verlagsgesellschaft mbH, Konstanz, München, 2015.
- [Ga84] Garfinkel, H.: Studies in ethnomethodology. Polity Press, Cambridge, 1984.
- [HT15] Hentze, M.; Tzafir, Y.: Freistunde. Doing Nothing All Day, 2015.
- [Jo18] Johnson, C. D. L.: Using virtual reality and 360-degree video in the religious studies classroom: An experiment, 2018; S. 228–241.
- [Kl13] Klemm, U.: Geschichte und Gegenwart Freier Demokratischer Schulen. Warte, bis der Schüler den ersten Schritt macht. In (Gürlevik, A.; Palentien, C.; Heyer, R. Hrsg.): Privatschulen versus staatliche Schulen. Springer VS, Wiesbaden, 2013; S. 115–136.
- [Le20] Leßner, T.: Regeln als Medium von Partizipation. Herstellung von Regeln zwischen institutionalisierter Schulversammlung und Alltag in Demokratischen Schulen. In (Munsch, C.; Müller, F. Hrsg.): Jenseits der Intention - Ethnografische Einblicke in Praktiken der Partizipation. Beltz Juventa, Weinheim, Basel, 2020, i.Dr.
- [MKH07] Martschinke, S.; Kopp, B.; Hallitzky, M.: Die Bedeutung von selbstreflexivem und forschendem Lernen in der Ausbildung für Lehramtsstudierende: PARadigma. Beiträge aus Forschung und Lehre aus dem Zentrum für Lehrerbildung und Fachdidaktik, 2007; S. 6–16.
- [Wi11] Wiesemann, J.: Ethnographische Forschung im Kontext von Schule. In (Moser, H.; Grunder, H.-U. Hrsg.): Forschung in der Lehrerbildung. Schneider-Verl. Hohengehren, Baltmannsweiler, 2011; S. 167–185.

## Virtual Reality in Health Care Education: A Study about the Effects of Presence on Acceptance and Knowledge Improvement among Health Care Students

Christian Plotzky<sup>1</sup>, Ulrike Lindwedel-Reime<sup>2</sup>, Alexander Bejan<sup>3</sup>, Peter König<sup>4</sup>,  
Christophe Kunze<sup>5</sup>

**Abstract:** With an ever-increasing need of skilled healthcare workers, efficient learning methods like Virtual Reality are becoming increasingly important. A VR simulation for endotracheal suction intervention was developed and tested.

The aim of this pilot study is to examine the VR simulation's acceptance and knowledge improvement. Furthermore, the effect of presence on acceptance and knowledge improvement was investigated.

A total of 51 students participated in the pilot study, using pre-post-test design. A modified Unified Theory of Acceptance and Use of Technology (UTAUT) and Igroup Presence Questionnaire (IPQ) were used. Correlation and regression analysis were performed. Pre- and post-tests showed a significant knowledge improvement ( $p < 0.001$ ). Correlation between presence and behavioural intention was highly positive ( $r = 0.52$ ,  $p < 0.001$ ).

Performance and effort expectancy are dominant in effect on behavioural intention of using VR learning simulation. The results indicate that a simulation, which conveys a higher sense of presence, is more likely to be accepted.

**Keywords:** Virtual Reality, Health Care, Education Technology, Acceptance, Presence, Knowledge Improvement.

### 1 Background and motivation

Like many other countries, the number of people in need of care in Germany has been rising steadily for decades [St18]. Since the number of personnel is not developing accordingly, a shortage of care workers is emerging [St18], [LF19]. This inevitably leads to a higher demand of nurses requiring more efficient education in large scale. In view of the tense personnel situation in health care, especially the practical instruction of students

---

<sup>1</sup> Hochschule Furtwangen, Institut Mensch, Technik und Teilhabe, Robert-Gerwig-Platz 1, 78120, Furtwangen, plan@hs-furtwangen.de

<sup>2</sup> Hochschule Furtwangen, Institut Mensch, Technik und Teilhabe, Robert-Gerwig-Platz 1, 78120, Furtwangen, liru@hs-furtwangen.de

<sup>3</sup> Hochschule Furtwangen, Institut Mensch, Technik und Teilhabe, Robert-Gerwig-Platz 1, 78120, Furtwangen, beja@hs-furtwangen.de

<sup>4</sup> Hochschule Furtwangen, Institut Mensch, Technik und Teilhabe, Robert-Gerwig-Platz 1, 78120, Furtwangen, koep@hs-furtwangen.de

<sup>5</sup> Hochschule Furtwangen, Institut Mensch, Technik und Teilhabe, Robert-Gerwig-Platz 1, 78120, Furtwangen, kuc@hs-furtwangen.de

is increasingly becoming a burden. Activities under instruction can therefore only be practiced to a lesser extent, if at all. This often results in a lack of readiness and skills among nurses, which has a negative impact on the quality of care and can lead to life-critical nursing errors [Ai12]. For this purpose, new learning methods such as VR must be studied.

Little is known about VR learning in nursing. So far, only few pilot studies have been conducted [We18], [BC19]. In the context of home ventilation as an exemplary field of application, endotracheal suction and ventilation have shown to be particularly stressful [LK18]. If nursing staff can perform suction carefully and without injury, relatives and patients feel safe [LE20]. Hence, a VR simulator for practicing the endotracheal suction procedure was developed. The aim of this paper is to estimate the target groups acceptance, and to validate knowledge improvement using VR in a pilot study. Furthermore, insights on how the sense of presence affects behavioural intention and knowledge improvement are acquired.

## **2 Theoretical framework**

### **2.1 Presence in VR and consequent potential for learning applications**

According to psychological learning theories like constructivism, experiential and situated learning, practical, interactive and lively experiences are highly beneficial for acquiring theory knowledge and practical skills in specialized field like nursing [Lo15]. Due to its immersive nature, VR has the capability of absorbing the user in a virtual environment creating an imprinting experience, hence leveraging the idea behind these learning theories. One construct to measure this experience is called presence. There are various definitions of presence, most of them being similar, for example “Presence is a state of consciousness, the (psychological) sense of being in the virtual environment.” [SW97] or “presence is the subjective sense of being in a virtual environment” [Sc03]. Based on learning theories, this subjective state should enhance the learning process for acquiring the desired knowledge and skills [S117].

One assessment tool for measuring presence is the Igroup Presence Questionnaire (IPQ) which divides presence in three subscales that are (I) Spatial Presence - the sense of being physically present in the VE; (II) Involvement - measuring the attention devoted to the VE and the involvement experienced; (III) Experienced Realism - measuring the subjective experience of realism in the VE [Sc03]. Due to its multi-dimensional nature and confirmed validity, IPQ was used as theoretical framework for measuring presence in this work.

## 2.2 VR in medical technology research and assessing knowledge improvement

VR simulators in medicine are widespread and there are even ISO-standards for validation [Pa18]. A survey of 31 studies with 2407 participants [Ky19] found evidence suggesting that VR improves post intervention knowledge and skills outcomes of health professionals when compared to traditional education or other types of digital education such as online or offline digital education. In the survey paper, knowledge improvement is defined as learner's factual or conceptual understanding measured using change between pre- and post-test scores. The definition was adopted for this paper.

## 2.3 Assessing technology acceptance

While knowledge improvement of a new learning tool is important, the tool also must be accepted and used by the target group. According to the unified theory of acceptance and use of technology (UTAUT) technology acceptance is defined as a process. A process that reaches from testing a technology, developing behavioural intention and eventually using the technology in the field [VMD03]. UTAUT is a validated model that combines eight technology acceptance models. The UTAUT model defines constructs, that causally affect behavioural intention. There are four key constructs: Performance expectancy, effort expectancy, social influence and facilitating conditions and constructs that have been identified to not significantly affect behavioural intention: Anxiety and self-efficacy. The constructs consist of multiple items, which are statements towards the technology. For example, one item for assessing performance expectancy would be "I would find the system useful in my job" [VMD03]. To apply UTAUT these items must be adapted to the setting and usage scenario of the technology being evaluated.

# 3 Material and methods

## 3.1 The VR simulation

The VR simulation has been developed at the Care and Technology Lab (Institut Mensch, Technik und Teilhabe), Hochschule Furtwangen. The simulation runs on the HTC Vive Head Mounted Display (HMD) with Vive controllers. The aim of the simulation is to execute the steps of endotracheal suction intervention, shown in **figure 1**, on a virtual home patient guided by audio voice records and text messages on a virtual tv screen. The steps followed by the learner reach from disinfecting hands at the beginning up to disposing the used equipment after the procedure. The order of steps follows the SOP of a German hospital.



Fig. 1: Screenshot collage of the VR simulation

The patient reacts in a realistic manner, looking at the learner, showing discomfort and coughing with a painful expression when the catheter is inserted in too deeply. Sound effects have been recorded from real medical equipment. After execution, the learner is presented an evaluation of his performance including time required for the suction procedure as well as number of faulty actions like contaminations. The simulation was developed iteratively and continuously checked by health care professionals and corrected accordingly. Moreover, a pilot study with nursing professionals who have had previous experience in endotracheal suction was conducted and usability of the simulation as well as correctness of the procedure were adapted based on the results.

### 3.2 Objectives

Due to the complexity of the endotracheal suction procedure, there is no standard tool yet for assessing performance in terms of skills (like OSCEs), which is why this study focused on knowledge improvement rather than skills improvement. Skill assessment will be a topic of further research. Since there has not been much research on VR in nursing education (see **chapter 1**), the conducted study's approach was based on a semi-explorative design, analysing many hypotheses, and revealing further research questions. The large number of hypotheses compared to the relatively low number of participants can lead to errors, however we do not draw final conclusions, but use the results to provide further insight on possible research topics. The first objective was to investigate if the UTAUT constructs in the special context of VR learning in nursing affect behavioural intention the same way as in general technological acceptance domains like bureau jobs.

Besides the four key constructs, anxiety and self-efficacy were included as these could play a role in the domain of acceptance of VR learning. This led to the following hypotheses:

**h1:** [a-f] has an effect on behavioural intention to use a VR simulation for learning.

**h1a:** Performance expectancy, **h1b:** Effort expectancy, **h1c:** Social influence, **h1d:** Facilitating conditions, **h1e:** Anxiety, **h1f:** Self efficacy

The second objective was to analyse how presence affects behavioural intention of using VR learning simulations. The presumption behind this idea was, that users who perceive a simulation as highly unrealistic, are more likely to not accept the simulation as a learning tool. Hence, user who feel more present should be more likely to accept the simulation as a learning tool.

**h2:** [a-d] has a positive effect on behavioural intention to use a VR simulation for learning.

**h2a:** Sense of presence, **h2b:** Spatial presence, **h2c:** Involvement, **h2d:** Experienced realism

The third objective was to investigate the effect of presence on knowledge improvement. Based on learning theories (see **chapter 2**) presence should have a positive effect on learning outcomes. This led to the following hypotheses:

**h3:** [a-d] in the VR learning simulation has a positive effect on knowledge improvement of the intervention.

**h3a:** Sense of presence **h3b:** Spatial presence, **h3c:** Involvement, **h3d:** Experienced realism

### **3.3 Study design and evaluation process**

In order to evaluate the hypotheses and gain further insight, 51 physiotherapy students were acquired for a study. Three VR setups separated by visual covers were installed in a large room. The students waited outside and filled out a knowledge pre-test. They were told that this test does not influence their grade in any way and were observed by an instructor to not cheat in the test. The students were asked to enter the room separately by the instructor whenever a VR setup was available to use. Next, they were given a standardised scripted explanation of how to use the system by one of three more instructors next to each VR setup. After this, they were requested to go through the simulation twice, in order to make sure they could learn the controls during the first run and focus on the task during the second run. Finally, they were asked to fill out the knowledge post-test and a UTAUT questionnaire as well as the IPQ.

Knowledge improvement (see **chapter 2**) was measured with the percentage decrease of error of the post-test compared to the pre-test. The tests were equal and contained 14 items, which were the steps of the intervention, that had to be put in the correct order. Error was calculated using the mean of absolute deviation from each item's put position by the learner compared to its correct position. Percentage decrease of error in the post-test compared to the pre-test was calculated and used as the score to measure knowledge improvement. Every student was taught the same knowledge about the intervention before the test. They learned from a theory lesson and an information sheet. To validate h1, a paired t-test was used on the students pre- and post-test scores.

In order to evaluate the hypotheses 2a - 2f, UTAUT (see **chapter 2**) was used. Items were translated into German and adapted to the context VR learning for students, trying to stay as close to the semantic meaning as possible. Internal consistency using Cronbach's alpha was measured to make sure constructs are reliable. To estimate effect of the constructs on behavioural intention partial least squared path modelling was performed as suggested by Venkatesh et al (2003). Calculations were performed with the PLS-PM library in R.

Presence was measured using the IPQ. The IPQ guide on how to calculate subscales was followed [Sc03]. In order to evaluate presence related hypotheses Pearson's correlation analysis was performed. The standard significance level of 0.05 was used throughout evaluation. Statistical calculations were performed using R.

## **4 Results**

### **4.1 Study sample and knowledge improvement**

Out of the 51 participants, 47 filled out the IPQ and UTAUT questionnaires, while 45 filled out the knowledge tests.



category	participants	gender		age			experience in	
subcategory	overall	m	f	18-25	26-30	>30	VR	suction
n	47	7	40	43	2	2	18	5
rate	100%	14.9%	85.1%	91.5%	4.2%	4.2%	38.3%	10.6%

Tab. 1: Study sample

Table 1 summarizes the study sample. The sample was relatively homogeneous consisting of students with health care profession, mostly at ages  $\leq 25$  (91.5%) and most being female (85.1%). 38.3% had previous experience in VR and 10.6% had previously performed endotracheal suction on a human patient. There was no significant difference in test performance between the group with and without previous VR experience.

As a requirement for further research questions, knowledge improvement through using the VR simulation had to be confirmed. Mean absolute error in the pre-test over all participants was 2.26 and 0.79 in the post-test. Mean percentage decrease of error was 65%. A paired t-test confirmed high significance ( $p < 0.001$ ).

#### 4.2 UTAUT constructs effect on acceptance

While four of the six UTAUT constructs reached at least a value of Cronbach's Alpha  $\alpha > 0.8$ , which indicates a construct with good internal consistency, *social influence* ( $\alpha = 0.64$ ) and *facilitating conditions* ( $\alpha = 0.37$ ) did not reach the  $\alpha \geq 0.7$  border which is considered unacceptable reliability. Thus, the two constructs had to be discarded for further conclusions leaving **h1c** and **h1d** unanswered.

construct	mean	effect ( $f^2$ )	t-value	p-value
Performance expectancy	4.4/5	0.34	2.33*	0.025
Effort expectancy	4.03/5	0.25	1.80**	0.008
Anxiety	1.74/5	-0.10	-0.83	0.41
Self-efficacy	3.36/5	0.11	0.88	0.38
Behavioural intention	3.8/5	-	-	-

Tab. 2: UTAUT constructs mean and effect on behavioural intention (\* $p < 0.05$ , \*\* $p < 0.01$ )

Table 2 shows the mean score of the UTAUT constructs as well as their effect on behavioural intention and its significance. Overall, behavioural intention of using VR for learning was high (3.8/5 (76%)). The constructs performance expectancy ( $f^2=0.34$ ,  $p=0.025$ ) and effort expectancy ( $f^2=0.25$ ,  $p=0.008$ ) had a significant effect with a medium to large effect size on behavioural intention suggesting that **h1a** and **h1b** are likely to be true.

### 4.3 Effects of presence

The scores of the presence subscales were as follows. The general sense of presence was relatively high (4.5/6), spatial presence (4.3/6) and involvement (3.9/6) were in the medium range, while the expected realism was rather low (2.5/6).

construct	Correlation (r)	t-value	p-value
sense of presence	0.52	4.07***	0.000
spatial presence	0.39	2.86*	0.01
involvement	0.26	1.79	0.08
experienced realism	0.34	2.43*	0.02

Tab. 3: Pearson's correlation between presence and behavioural intention (\*\*\* $p<0.001$ , \* $p<0.05$ )

**Table 3** shows Pearson's correlations between the presence subscales and behavioural intention. Correlation between sense of presence and behavioural intention was highly significant ( $p<0.001$ ) and according to Cohen also has a large effect ( $r>0.5$ ). Correlation between spatial presence and behavioural intention, as well as experienced realism were also significant with a medium effect. Thus, **h2a**, **h2b** and **h2d** were confirmed. The effect of involvement was not significant and had only a low effect size, which lead to **h2c** being rejected.

construct	Correlation (r)	t-value	p-value
sense of presence	-0,28	-1,89	0,07
spatial presence	-0,33	-2,31*	0,03
involvement	-0,15	-1,03	0,31
experienced realism	-0,16	-1,07	0,29

Table 4: Pearson's correlation between presence and knowledge improvement (\* $p < 0,05$ )

Likewise, **table 4** shows correlations between the presence subscales and knowledge improvement. Correlation between spatial presence and knowledge improvement was negative with medium effect size ( $r = -0.33$   $p = 0.03$ ). **h3b** was not confirmed, due to the negative effect. The other presence subscales did not correlate significantly ( $p > 0.05$ ) which means **h3a**, **h3c** and **h3d** were not confirmed either.

## 5 Discussion

Like the UTAUT model suggests, performance expectancy and effort expectancy had a significant effect on behavioural intention. Although some participants showed fear during execution of the simulation and had to cancel it, anxiety did not have a significant effect and neither did self-efficacy. Due to their insufficient reliability, no statement about social influence and facilitating conditions can be made. Overall, technology acceptance in VR learning does not seem to differ from other domains considering the UTAUT constructs. However, the number of participants were small and further research needs to be done to verify this, especially regarding the anxiety construct, since our subjective observations contrasted this.

Sense of presence had a medium to strong size positive effect on behavioural intention that was highly significant (see **chapter 4**). A higher sense of presence in a simulation could lead to a higher behavioural intention by learners. This seems plausible, since a simulation that absorbs the learner into a virtual training scenario and feels subjectively more realistic, should be more likely to be accepted and used.

Out of the presence metrics, only spatial presence correlated significantly ( $p < 0.05$ ) with knowledge improvement. It seems likely that a user who feels more spatially present in the virtual training scenario would have a more imprinting experience and could therefore memorize more of it. However, against this presumption the correlation was negative.

Regarding the effects of presence on learning outcomes, there is no consensus in literature. While [SGL19] and [Ch18] found a positive correlation between presence and learning outcomes, [Ba12] found a negative correlation (like this research) and [Pe09] as well as [LPF19] found no significant correlation at all. There could be various reasons for this discrepancy. The first reason could be that there are various survey instruments for assessing presence and learning outcomes which could lead to not measuring the same constructs. Another more complex explanation includes viewing presence from two different perspectives. First, presence can be seen from user perspective suggesting that each user has a different sense of presence towards one simulation. Second, presence can be seen from simulation perspective suggesting that the same user feels different degrees of presence in different simulations. This leads to two different study designs. In the first study design, differences in the degree to which user perceive presence in the same simulation are measured. In the second study design different simulations are compared by having the same users participate in multiple simulations and measure the differences in degrees of presence which simulations can convey. On the one hand, users who generally feel more present in VR simulations than other users might focus more on the environment than the actual task. On the other hand, simulations that generally convey a higher sense of presence could lead to more imprinting experiences and therefore positively affect knowledge improvement.

### **5.1 Limitations**

There are some flaws in the study design. On the one hand, the VR simulation aims to teach nursing skills. The participants, however, were physiotherapy students as there were no larger nursing student groups available to the authors for testing. Additionally, previous experience in suction is likely to influence performance. On the other hand, the large amount of effects looked upon with a relatively small amount of study participants could lead to statistical errors. Another problem is that no split of groups was done. Participants all ran the same simulation which lead to making assumptions based on general effect size classes and correlation of constructs rather than comparing effect sizes.

### **5.2 Conclusion**

Our pilot study analysed various aspects of VR learning in the health care education sector. In the study, performance expectancy and effort expectancy had a significant effect on behavioural intention in the target group while self-efficacy and anxiety did not. This reflects the results from the creators of UTAUT [VMD03]. We could not confirm the effects of social influence and facilitating conditions due to unreliability of our constructs. Moreover, our results suggest that the sense of presence could have a positive effect with large size on behavioural intention of health care VR learning simulations. In further research, the UTAUT based items must be improved in order to more reliably gain insight,

especially about the anxiety construct, which according to our observations seemed to play a role. Moreover, our detected effect of presence must be confirmed through further studies.

Finally, we found that there was a significant negative correlation between spatial presence and knowledge improvement. We explained a possible reason for the discrepancy of the measured effects of presence on learning outcomes in different studies. There is still no consensus about the effect of presence on learning and further research needs to investigate the reasons for this.

## Bibliography

- [Ai12] Aiken, L. H. et al.: Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. In *BMJ (Clinical research ed.)*, 2012, 344; e1717.
- [Ba12] Bailey, J. et al.: Presence and Memory: Immersive Virtual Reality Effects on Cued Recall: Proceedings of the International Society for Presence Research Annual Conference, 2012.
- [BC19] Bayram, S. B.; Caliskan, N.: Effect of a game-based virtual reality phone application on tracheostomy care education for nursing students: A randomized controlled trial. In *Nurse education today*, 2019, 79; pp. 25–31.
- [Ch18] Cho, Y.: How Spatial Presence in VR Affects Memory Retention and Motivation on Second Language Learning: A Comparison of Desktop and Immersive VR-Based Learning, 2018.
- [Ky19] Kyaw, B. M. et al.: Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. In *Journal of medical Internet research*, 2019, 21; e12959.
- [LE20] Lehmann, Y.; Ewers, M.: Sicherheit in der häuslichen Intensivversorgung beatmeter Patienten aus Sicht professioneller Akteure. In *Gesundheitswesen (Bundesverband der Ärzte des Öffentlichen Gesundheitsdienstes (Germany))*, 2020, 82; pp. 75–81.
- [LF19] Lindner, S.; Fleischer, N.: Arbeitsmarktsituation im Pflegebereich, Nürnberg, 2019.
- [LK18] Lindwedel-Reime, U.; König, P.: Wahrgenommene Belastungen professionell Pflegenden in der außerklinischen Beatmungspflege. Welchen Einfluss hat die Technik? In *Zukunft der Pflege - Tagungsband der 1. Clusterkonferenz*, 2018, 2018; pp. 142–146.
- [Lo15] Loke, S.-K.: How do virtual world experiences bring about learning? A critical review of theories. In *Australasian Journal of Educational Technology*, 2015, 31.

- [LPF19] Lerner, D.; Pranghofer, J.; Franke, A.: Der Einfluss des Präsenzerlebens auf die Lern- und Trainingseffekte in einer Virtual-Reality Simulationsumgebung. Fraunhofer-Gesellschaft, 2019.
- [Pa18] Pantelidis, P. et al.: Virtual and Augmented Reality in Medical Education. In (Tsoulfas, G. Ed.): Medical and Surgical Education - Past, Present and Future. InTech, 2018.
- [Pe09] Persky, S. et al.: Presence relates to distinct outcomes in two virtual environments employing different learning modalities. In *Cyberpsychology & behavior the impact of the Internet, multimedia and virtual reality on behavior and society*, 2009, 12; pp. 263–268.
- [Sc03] Schubert, T. W.: The sense of presence in virtual environments. In *Zeitschrift für Medienpsychologie*, 2003, 15; pp. 69–71.
- [SGL19] Selzer, M. N.; Gazcon, N. F.; Larrea, M. L.: Effects of virtual presence and learning outcome using low-end virtual reality systems. In *Displays*, 2019, 59; pp. 9–15.
- [SI17] Slater, M.: Implicit Learning Through Embodiment in Immersive Virtual Reality. In (Liu, D. et al. Eds.): *Virtual, Augmented, and Mixed Realities in Education*. Springer Singapore, Singapore, 2017; pp. 19–33.
- [St18] Statistisches Bundesamt (Destatis): *Pflegestatistik: Pflege im Rahmen der Pflegeversicherung. Deutschlandergebnisse 2017*, 2018.
- [SW97] Slater, M.; Wilbur, S.: A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. In *Presence: Teleoperators and Virtual Environments*, 1997, 6; pp. 603–616.
- [VMD03] Venkatesh; Morris; Davis: User Acceptance of Information Technology: Toward a Unified View. In *MIS Quarterly*, 2003, 27; p. 425.
- [We18] Weiss, S. et al.: Applications of Immersive Virtual Reality in Nursing Education - A Review: 1. Clusterkonferenz Zukunft der Pflege -- Innovative Technologien für die Praxis, 2018.

## DigiCare - Ein intelligentes Lehr-/Lernsystem für Gesundheitsberufe

Nikolaj T. Graf von Malotky <sup>1</sup>, Alke Martens<sup>2</sup>

**Abstract:** DigiCare ist ein Projekt, das vielfältige Ansätze kombiniert um Studiengänge des Gesundheitsmanagements zu digitalisieren und damit den Anforderungen einer heterogenen Zielgruppe gerecht zu werden. Die im Projekt verfolgten Ansätze kombinieren klassische Vorlesungsaufzeichnungen, Lernerannotation bei Live Veranstaltungen mit intelligenter Analyse der Eingaben, sowie intelligentes fallbasiertes Tutoring mit natürlichsprachlichen Dialogen.

**Keywords:** Intelligentes Tutoring, Dialog, fallbasiertes Lernen, Gesundheitsmanagement

### 1 Einleitung

Die Gesundheitswirtschaft steuert auf eine Situation zu, in der es einen eklatanten Mangel an Fachkräften geben wird [NeHo17]. Auch in diesem Bereich, der traditionell eher auf zwischenmenschliche Kontakte setzt, als auf digitale Techniken, halten technische Innovationen zunehmend Einzug. Es geht vor allem darum, den Einsatz von Pflegekräften effizienter zu machen, indem Prozesse digital dokumentiert, schnell zugreifbar abgelegt und strukturiert verfügbar gemacht werden. Interessant ist dabei auch, das aktuelle Lohngefälle zwischen technischen und sozialen Berufen [Web18]. Die Hochschule Neubrandenburg und auch die Universität Rostock streben daher an, Fachpersonen mit hochschulischen Abschlüssen und breiter Qualifikation in den Bereichen Gesundheit, Pflege und Management auszubilden. Daraus resultierten für das Projekt zwei Entwicklungslinien. In der ersten Entwicklungslinie wird das Lehrmaterial selbst digital aufbereitet und in verschiedenen Formaten zur Verfügung gestellt (siehe hierzu [DrMa19], [Ni19]). Studienbegleitend wird ein fallbasiertes intelligentes Tutoringsystem (ITS) mit der Möglichkeit der Dialogführung mit virtuellen Patienten entwickelt und in den Studiengang ab einem frühen Semester integriert. Im Fokus dieses Papers steht diese zweite Entwicklungslinie. Generell ist der Ausbildungskontext Gesundheit und Pflege eng verbunden mit praktischen Fähigkeiten und Fertigkeiten. Vor allem die pflegerischen Handlungskompetenzen lassen sich eigentlich nur in der praktischen Anwendung vertiefen und ausbauen. In der klassischen hochschulischen Ausbildung fehlt die Praxisnähe. Eine Integration von Pflegetrainings auf der Basis von

---

<sup>1</sup> Universität Rostock, Institut für Informatik, Albert-Einstein-Str. 22, 18059 Rostock, Deutschland  
nikolaj.graf\_von\_malotky@uni-rostock.de ,  <https://orcid.org/0000-0000-0000-0000>

<sup>2</sup> Universität Rostock, Institut für Informatik, Albert-Einstein-Str. 22, 18059 Rostock, Deutschland  
alke.martens@uni-rostock.de

Praktika oder Seminaren ist kaum zu realisieren, weil nicht genügend Trainingspatienten zur Verfügung stehen und das Training am Kranken- oder Pflegebett auch unter Berücksichtigung von ethischen Prinzipien kaum als Bestandteil eines Studiums zu realisieren ist. Gerade bei zunehmender Anzahl Studierender kann oft der Praxisbezug nur durch die (nicht interaktive) Präsentation von Lehrfällen hergestellt werden. Dieser Praxisbezug findet dann kognitiv nur auf der Metaebene statt: die tatsächliche Konfrontation mit Situationen, die als beängstigend oder verunsichernd erlebt werden, können auf diese Weise nicht trainiert werden. Ein patientennahes, in der Didaktik auch als fallbasiert [ZuHa08] bezeichnetes Training erfolgt im Studium kaum. Hier kann ein fallbasiertes, dialogorientiertes Intelligentes Tutoring System Abhilfe schaffen. Das im Projekt umgesetzte fallbasierte intelligente Tutoring wird in den nachfolgenden Abschnitten vertieft dargestellt.

## **2 Fallbasiertes intelligentes Tutoring**

Seit einigen Jahren gibt es Erfahrungen im Bereich der medizinischen Ausbildung, die belegen, dass fallbasiertes Lernen von Handlungsschritten durch computerbasierte Anwendungen (sogenannte virtuelle Patienten oder virtuelle Behandlungssituationen) zielgerichtet unterstützt werden kann (siehe z.B. [Nikolaj], [Ma04], [PaRoZa12]). Das System der Wahl ist das Intelligente Tutoring System, das mit breiten Expertensystembasen von Fachwissen (in diesem Fall medizinisches und pflegerisches Wissen) arbeitet (siehe z.B. [HarrerMartens] und auch [NiMa20]) und damit auch interaktive Dialoge ermöglicht (z.B. [Ca70], [RuMo12], [Ri99], [NiMa20]).

Die klassische ITS Architektur besteht wenigstens aus den folgenden Komponenten: User Interface Component - die Komponente der Benutzungsschnittstelle; Domain Knowledge Component - die Komponente, die früher auch als Expertensystemschnittstelle bezeichnet wurde und in der das Domainwissen verankert ist; Pedagogical Knowledge Component(s) - die Komponente(n), in der das Wissen um die didaktische Strukturierung des Lehrmaterials gehalten wird und schließlich die Student Knowledge Component - die Komponente, in der das Wissen des Lernenden und das Wissen über den Lernenden gehalten wird. Verschiedene Formen heterogener Nutzung wurden unter anderem in [Ma04] und [NiMa20] beschrieben und werden im Folgenden nicht weiter ausgeführt. Das Hauptproblem der klassischen Komponenten besteht darin, dass Funktionalität und intendierte Nutzung oft vermischt genutzt werden (siehe z.B. [Ma04], [NiMa20]). DigiCare bedient sich hier einer neuen generischen ITS Architektur, die eine funktionsorientierte Sichtweise einnimmt und als Framework auch anderen ITS Entwicklungen zur Verfügung steht. Die funktionsorientierte Sichtweise, die auf einer breiten Analyse bestehender ITS Architekturen basiert [NiMa20], mündet in drei Funktionsbereichen. Diese drei Funktionsbereiche werden in einer Schichtenarchitektur umgesetzt. In jeder Schicht wird dann separat ausgebaut, welche Funktionen in der Schicht aus Sicht eines ITS - oder sogar eines beliebigen anderen E-Learning Systems - umgesetzt werden können. Von implementationsspezifischen Details



wird in der generischen Basisarchitektur abgesehen. Die drei Funktionsbereiche sind: User Facing Components: dies entspricht Interaction Layer in der Schichtenarchitektur; Functionality Components: dies entspricht Functionality Layer in der Schichtenarchitektur und Database Facing Components: dies entspricht Data Layer in der Schichtenarchitektur.

Es gibt neben den drei Funktionsschichten noch zwei weitere Schichten: den Temporary Layer und den Persistence Layer. In diesen Schichten tritt eine neue Sicht auf ITS zutage: die Temporary Layer definiert softwareseitig, welche Arten von Sichten für den Lernenden zur Verfügung stehen. Der Persistency Layer bestimmt, welche Gruppen von Daten in dem ITS vorliegen und wie diese Daten gespeichert werden sollen. Die Gruppierung erfolgt aufgrund von Ähnlichkeiten und stellt sicher, dass Lernerdaten von Expertenwissensdaten unterschieden werden können. Die Unterscheidung hat keine Performancegründe, sondern ist in der Struktur der Architektur begründet. Die klassischen Komponenten sind im Persistence Layer abgebildet: Domain Knowledge Component - das Expertenwissensmodell; die Pedagogic Component - das pädagogische Wissen, sowie die Student Knowledge Component - das Lernermodell bzw. die Komponente, die dieses enthält. Zusätzlich gibt es noch eine Program Knowledge Component, die Ausführungs- und Adaptionen funktionalität des ITS bereithält.

Eine Ausführungskomponente erweitert in der klassischen ITS Architektur. Dies hatte historisch dazu geführt, dass mal die eine und mal die andere Komponente die Ausführung von Adaptionen funktionalitäten und anderen Berechnungen zur Laufzeit ausführte (z.B. Antwortgenerierung, Lerneranalyse etc) (siehe [Ma04]). Die Einführung einer Extrakomponente (der Program Knowledge Component) löst dieses Problem in der generalisierten Schichtenarchitektur. Die Komponenten aus der Persistence Layer werden im Functionality Layer zusammengefügt zur ITS Core Functionality und zu weiteren Program Functionalities. Dies entspricht der Trennung im Persistence Layer: die Kern funktionalitäten eines ITS in den klassischen Komponenten plus der Zusatzkomponente in Program Functionalities. Im Temporary Layer werden schließlich die Teaching Scenes, also Lehrszenarios, und weitere Szenarios aufbereitet. Diese werden an den Lerner angepasst im Interaction Layer angezeigt. Die Trennung zwischen Teaching Scenes und Additional Scenes entstammt der Überlegung, dass in den meisten ITS eine Kombination von eigentlichem für die Lehre aufgearbeiteten Material mit allgemeinem, nicht aufgearbeiteten Material stattfindet. Dass die Generalisierte ITS Architektur noch viel mächtiger ist, zeigt die folgende Abbildung 1, in der der aktuelle Stand mit allen verfügbaren Schichten und Komponenten zu sehen ist. Eine genauere Beschreibung ist in [NiMa20] zu finden. Innerhalb der Components in den Layer wird nun die eigentliche Funktionalität eines ITS konzeptionell ausgebaut und programmiert. Mit mehr Details wird dies beschrieben in [NiMa20].

Im Projekt DigiCare liegt der Fokus auf zwei wesentlichen Gestaltungselementen für die Umsetzung und Nutzung der Fünf-Schichten Architektur. Dies sind die didaktische Entscheidung für fallbasiertes Lernen in einem medizinischen Bereich, der jedoch nicht auf diagnostisches Schließen fokussiert, sondern auf die Interaktion zwischen dem

Lernenden und den angebotenen Informationen. Der Lerner lernt eher entscheidungsbasiert – dies bedeutet, dass die Adaptionfunktionalität des Gesamtsystems stark sein muss, da die Entscheidungen des Lerners fortwährend die Entwicklung des Lehrfalls beeinflussen können soll. Zudem ist die wichtige Funktionalität des Dialogs mit einem virtuellen Charakter im System umzusetzen.

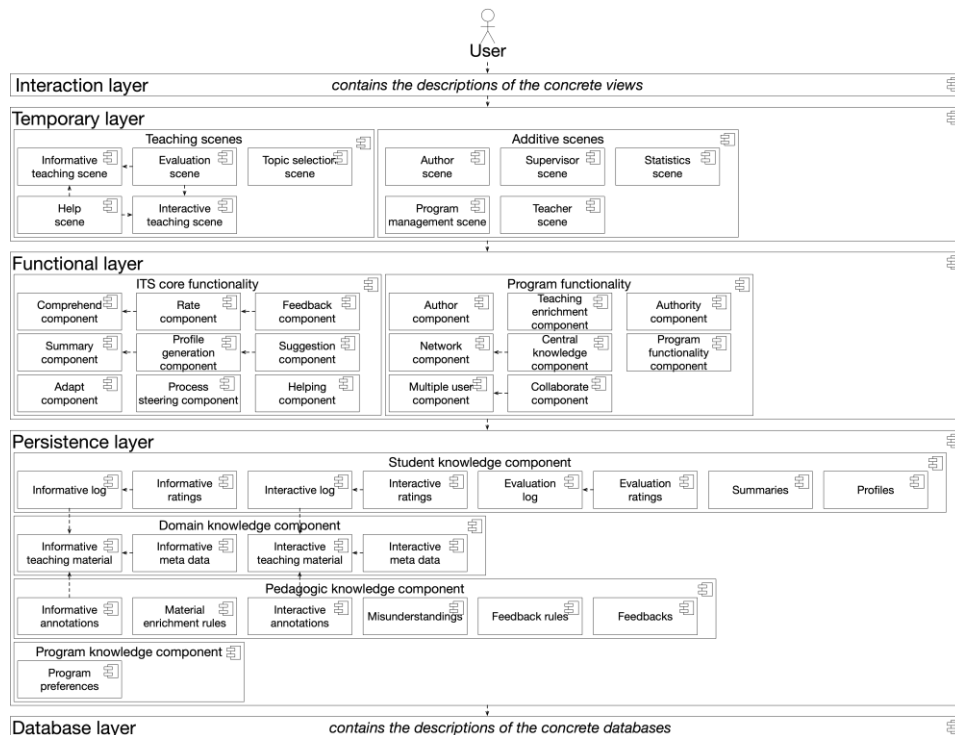


Abb 1: Generelle Architektur des ITS bei der Umsetzung in DigiCare (aus: [vMM20])

Das resultierende System mag zunächst komplex erscheinen, doch ist es vom Aufbau her an die obige Fünf-Schichten Architektur angelehnt. Die unterste Schicht wird aus den Datenstrukturen gebildet, die durch Datenbanken realisiert sind. Der Zugriff auf diese Datenbestände erfolgt in der Regel lesen – allein bei dem Modul Student Progress muss eine Komponente in der Lage sein, selbständig Einträge vorzunehmen. In der darüberliegenden Persistency Layer ändert sich im Vergleich zur oben aufgeführten Architektur ebenfalls nichts. Alle genannten Bestandteile werden in Form von Controllern umgesetzt. Die Bezeichnung als Controller kommt daher, dass im Projekt entschieden wurde, eine Model-View-Controller Architektur aus Softwaresicht zu realisieren. Die Fünf-Schichten Architektur unterstützt diesen Ansatz. Der erwähnte Chatbot wird als Teil des Functionality Layers umgesetzt. Das System muss dabei in der Lage sein, die Lernereingabe (Freitext) zu analysieren, einen bestimmten Dialogzustand zu analysieren und herauszuarbeiten und basierend auf diesen beiden Informationen eine

Antwort zu generieren. Die Antwort wird dann an den Interaction Layer weitergegeben und dann im Interface dem Lerner angezeigt. Im modellierten Lehrfall könnte dies beispielsweise ein Dialog mit einem demenzkranken Patienten sein. Der Interaction Layer wird in Form einer REST API Backend ausprogrammiert. Hier werden sowohl die interaktiven Lehrfallaufgaben (Interactiv Task Controller) als auch die Chats aus Interaktionssicht umgesetzt. In Fall von DigiCare gab es zudem noch die Anforderung, für ein Webinterface zu entwickeln. Daher liegt die Fünf-Schichten-Architektur auf einem Webserver, während der User in einem Webclient arbeitet (auf einem Browser seiner Wahl). Der Presentation Layer befindet sich entsprechend auf dem Webserver - hier finden alle Vorgänge in der Interaktion mit dem Lerner ihre Umsetzung: Login, Auswahl von Lehrfällen und interaktiven Aufgaben, als auch der Chat.

### **3 Zusammenfassung und Ausblick**

Digitales Lehren und Lernen in Studiengängen des Gesundheits- und Pflegemanagements ist eine Grundanforderung, die bereits vor der Corona-Krise bestanden hatte. Im Rahmen des Projektes DigiCare werden daher derzeit zwei verschiedene Foki bearbeitet: die Digitalisierung von klassischen Lehrveranstaltungen und die Aufarbeitung von Lehrinhalten für die digitale Verwendbarkeit und die Integration von fallbasiertem Lernen in Form eines ITS. Der vorliegende Artikel ist vor allem dem zweiten Aspekt gewidmet. In diesem Artikel wurde beschrieben, welcher Weg für die Realisierung der Basisarchitektur des Intelligenten Tutoring Systems vorgenommen wurde. Neu bei diesem Ansatz ist, dass auf Basis einer generalisierten Fünf-Schichten Architektur entwickelt wurde. Für die Fünf-Schichten Architektur ist das DigiCare ITS ein weiterer „Proof of Concept“, der gezeigt hat, dass der Ansatz einer generalisierten Architektur die Entwicklung eines ITS von der Entwurfsphase bis zur Implementierungsphase drastisch beschleunigen kann und hilft, etliche Desingentscheidungen differenzierter zur durchdenken, als dies auf Basis der bisherigen Architekturen möglich war.

### **Danksagung**

Unser Dank gilt dem gesamten Projektteam von DigiCare, vor allem unseren Projektpartner aus dem medizinischen Bereich, die uns mit viel Engagement bei der Entwicklung der Lehrfälle zur Seite stehen, und natürlich unseren Förderern. Das Verbundprojekt „DigiCare“ wird durch den Europäischen Sozialfonds (ESF) mit dem Förderkennzeichen ESF/14-BM-A55-0018/19 und dem Ministerium für Bildung, Wissenschaft und Kultur des Landes Mecklenburg-Vorpommern gefördert.

## Literaturverzeichnis

- [Ca70] Carbonell, J.R.: AI in CAI: An Artificial Intelligence Approach to Computer Assisted Instruction. *IEEE Transactions on Man-Machine-Systems*, 11(04):190–202, 1970
- [DM19] Drews, P.; Martens, A.: Juniorstudium - Study Digital while Going to School. In (M. Herzog, Z. Kubincova, H. Peng M. Temperini, Hrsg.): *Advances in Web-Based Learning -ICWL2019*. Springer LNCS, Virtual, 2019.
- [JHM08] J. Zumbach; Haider, K.; Mandl, H.: Fallbasiertes Lernen: Theoretischer Hintergrund und praktische Anwendung. In: *Pädagogische Psychologie in Theorie und Praxis. Einfallbasiertes Lehrbuch*. Hogrefe, Göttingen, S. 1–11, 2008.
- [Ma04] Martens, A.: Ein Tutoring Prozess Modell für fallbasierte intelligente Tutoring Systeme. DISKI 281, Verlag infix, AKA, Heidelberg, Berlin, 2004.
- [MH10] Martens, A.; Harrer, A.: Intelligent Tutoring Systems Architecture Rebuilt — A Pattern Approach. In: *Intelligent Tutoring Systems in eLearning Environments*. Information Science Publishing, Idea Group Inc., S. 70–86, 2010.
- [Ne17] Neldner, Thomas; Hofmann, Esther; Richter, Tobias; u. a.: *Entwicklung der Angebotsstruktur, der Beschäftigung sowie des Fachkräftebedarfs im nichtärztlichen Bereich der Gesundheitswirtschaft*. bmwi, Berlin, Darmstadt, Tübingen, 2017.
- [Ni19] Nicolay, R.; von Malotky, N.T. Graf; Auge, T.; Martens, A.: Autonomous Se-mantic Structuring of Lecture Topics - Synthesis of Knowledge Models, DOI:10.5220/0006367903490355. In (Uhomoihi, J., Hrsg.): *9th International Conference on Computer Supported Education*. Scitepress, Hierarclion, Greece, S. 349–355, 2019.
- [PRZ12] Paviotti, G.; Rossi, P.G.; Zarka, D.: *Intelligent Tutoring Systems: An Overview*. [http://blog.unimc.it/itutor/files/2012/06/Intelligent\\_Tutoring\\_Systems\\_An\\_Overview\\_2012.pdf](http://blog.unimc.it/itutor/files/2012/06/Intelligent_Tutoring_Systems_An_Overview_2012.pdf), last visited 2012., verified 06/2020
- [Re20] Redaktionsteam, *Praktischer Arzt, Was verdient ein Altenpfleger*. <https://www.praktischerarzt.de/medizinische-berufe/altenpfleger-gehalt/>, zuletzt besucht Juni 2020.
- [RJ99] Rickel, J.; Johnson, W.L.: Animated Agents for Procedural Training in Virtual Training: Perception, Cognition and Motor Control. *Applied Artificial Intelligence Journal*, 0(0):343–382, 1999.
- [Ru12] Rus, V.; Moldovan, C.; Graesser, A.; Niraula, N.: Automated discovery of speech act categories in educational games. In: (Yacef et al.) *Proceedings of the 5th Conference on Educational Data Mining*, Chania Greece, 2012. online S. 25–32.
- [vMM19] von Malotky, N.T. Graf; Martens, A.: A comparable foundation for ITS research loop and subfields, DOI: 10.21125/edulearn.2019.0425. In (Chova, L. Gómez; Martínez, A. López; Torres, I. Candel, Hrsg.): *911th International Conference on Education and New Learning Technologies*. IATED, Palma, Spain, S. 1402–1409, 2019.
- [vMM20] von Malotky, N.T. Graf; Martens, A.: General ITS software architecture and framework. In (Sgouropoulou, C., Hrsg.): *ITS Intelligent Tutoring Systems Conference 2020*. Springer, Virtual, 2020.

## Touch-Aktionen beim digitalen Lernen: Steigerung der Performance des visuellen Gedächtnisses durch aktive Reizverarbeitung

Thiemo Leonhardt <sup>1</sup>, Gregor Damnik<sup>2</sup> und Nadine Bergner <sup>3</sup>

**Abstract:** Das Wahrnehmen und Speichern von neuen Eindrücken und Erfahrungen aus der Umwelt ist eine notwendige Basis für einen Großteil an Denk- und Lernvorgängen. Entsprechend gut ist der Mensch darin, Reize aus seiner Umwelt aufzunehmen und in das Langzeitgedächtnis zu transferieren. Bisherige Experimente haben gezeigt, dass diese Behaltensleistung unter bestimmten Bedingungen optimiert werden kann. So können Reize in Form von Alltagsgegenständen dann gut memoriert werden, wenn sie nicht nur passiv betrachtet, sondern aktiv beispielsweise durch Betasten verarbeitet werden. Die vorliegende Studie überträgt diese Befundlage auf das Lernen mit digitalen Endgeräten mit Touch-Eingabeoptionen. Dazu haben 46 Personen Gegenstände am Bildschirm entweder nur betrachtet oder per Dragfunktion aktiv bewegt. Die Ergebnisse dieses Experiments belegen, dass der zusätzliche haptische Reiz die Behaltensleistung im visuellen Gedächtnis signifikant steigert und dass die Verwendung der Touch-Funktion damit neben besserer Usability auch einen Effekt auf den Lernprozess aufweisen kann.

**Keywords:** Touch-Interaktion, Touch-Displays, digitale Lernanwendungen, Gedächtnis, Behaltensleistung

### 1 Einleitung


Der Mensch ist jeden Augenblick seines Lebens vielen gleichzeitigen Wahrnehmungserfahrungen ausgesetzt. Deshalb scheint es fraglich, ob der Mensch in der Lage ist, alle diese Wahrnehmungserfahrungen im Langzeitgedächtnis abzulegen. Erkenntnisse im Bereich des visuellen Langzeitgedächtnisses zeigen, dass der Mensch fähig ist, eine sehr große Anzahl an Wahrnehmungserfahrungen zu speichern [St73]. Dabei werden nicht nur die groben Formen oder Kernelemente behalten, sondern auch spezielle Details [Br08] [Ko10] [VM07]. Ein interessanter Punkt hierbei ist, dass eine Lern- oder Behaltensabsicht nicht Voraussetzung für die Übernahme in das visuelle Langzeitgedächtnis ist, sondern diese Speicherung ein Ergebnis der visuellen Wahrnehmung darstellt [CH05] [KRS17] [WHZ05].

---

<sup>1</sup> Technische Universität Dresden, Professur für Didaktik der Informatik, 01062 Dresden,

thiemo.leonhardt@tu-dresden.de,  <https://orcid.org/0000-0003-4725-9776>

<sup>2</sup> Technische Universität Dresden, Professur für Didaktik der Informatik, 01062 Dresden, gregor.damnik@tu-dresden.de

<sup>3</sup> Technische Universität Dresden, Professur für Didaktik der Informatik, 01062 Dresden, nadine.bergner@tu-dresden.de  <https://orcid.org/0000-0003-3527-3204>

Eine hohe visuelle und haptische Erinnerungsleistung von Alltagsgegenständen bei Menschen ist selbst beim unbewussten Einprägen von Objekten deutlich erkennbar [HK18]. Das Überführen des Gelernten in das Langzeitgedächtnis scheint ebenfalls vorzuliegen. Die Kombination aus visuellen sowie haptischen Reizen bildet demnach einen vielversprechenden Ansatz diesen Effekt auch bei digitalen Lernanwendungen zu untersuchen.

Aus dieser Überlegung folgt die Fragestellung, ob eine minimale Haptik durch eine Touch-Interaktion einen Mehrwert beim Lernen gegenüber einem rein passiven Input durch Betrachtung bietet. Mit diesem Vorgehen werden die Forschungsergebnisse und Erkenntnisse aus der Mensch-Computer-Interaktion erweitert und der Aspekt des Lernens stärker in den Vordergrund bei der Entwicklung des Designs insbesondere den Interaktionsmöglichkeiten von Lernanwendungen gerückt. Die hier beschriebene Studie ist dazu ein erster Baustein.

## 2 Experiment

Die Fragestellung ist, ob neben dem visuellen Reiz beim Betrachten eines Gegenstands (Bildes) ein zusätzlicher für alle Gegenstände identischer haptischer Reiz die Behaltensleistung in Bezug auf das visuelle Gedächtnis signifikant steigern kann.

Abgeleitet aus der Fragestellung, ob eine minimale Haptik durch eine Touch-Interaktion einen Mehrwert beim Lernen gegenüber einem rein passiven Input durch Betrachtung bietet, werden die Nullhypothese und die Alternativhypothese gebildet. Als Nullhypothese wird angenommen, dass die Performance der Gruppe *view* größer oder gleich der Performance der Gruppe *draggable* ist.

$$H_0 : \text{Performance}_{\text{view}} \geq \text{Performance}_{\text{draggable}} \quad (1)$$

Als Alternativhypothese wird angenommen, dass die Performance der Gruppe *draggable* größer ist als die Performance der Gruppe *view*.

$$H_1 : \text{Performance}_{\text{draggable}} > \text{Performance}_{\text{view}} \quad (2)$$

### 2.1 Stichprobe

In vergleichbaren Studien zur Untersuchung des visuellen Langzeitgedächtnisses wurde bei [Br08] ein Stichprobenumfang von 14, bei [VM07] von 24 und bei [HK18] von 26 sowie 48 erreicht. Zur Vergleichbarkeit und zum Erhalt eines ausreichenden Stichprobenumfangs wurde für die hier vorliegende Studie ein Stichprobenumfang von  $N > 40$  angestrebt. Insgesamt haben 46 Personen teilgenommen, wobei Daten von zwei Personen aufgrund technischer Fehler nicht vollständig aufgezeichnet wurden. Der gültige Stichprobenumfang besteht demnach aus 44 Personen (Alter:  $M = 34.8 \text{ Jahre}$ ,  $SD = 10.6$ ,

*Mdn* = 33). Das Experiment wurde im Open Science Framework<sup>4</sup> registriert und die Daten sowie Testapplikationen können dort eingesehen werden.

## 2.2 Forschungsdesign

Das Forschungsdesign verfolgt insgesamt den Ansatz, dass den Teilnehmenden im Vorfeld nicht kommuniziert wird, dass eine Behaltensleistung gemessen wird. Ähnlich wie in der Studie von [HK18] wird den Teilnehmenden erläutert, dass das Design der Gegenstände beurteilt werden soll. Das Stimulusset besteht aus Photographien von 120 Paaren von unterschiedlichen Alltagsgegenständen verschiedener Kategorien. Jedes Paar von Objekten besteht aus zwei Vertretern derselben grundlegenden Kategorie (z. B. Scheren). Die Objekte eines Paares unterscheiden sich dabei in mehreren Eigenschaften zum Beispiel der Farbe, der Größe und dem Design. Die Photographien der Objekte wurde der Studie aus [HK18] entnommen, das Set kann beim Open Science Framework angerufen werden.

Die Teilnehmenden wurden automatisch nach den bis zu diesem Zeitpunkt gespeicherten Datensätzen der Gruppe *view* oder *draggable* zugeordnet, um eine Gleichverteilung zu erreichen. Die Teilnehmenden wurden per Mail oder Chat instruiert, etwa 20 Minuten konzentriert an dieser Studie zu partizipieren. Dazu musste ein eigenes Tablet mit Touch-Funktionalität benutzt werden. Es musste auf dem Gerät eine Webseite aufgerufen werden, über die das Experiment ablief. Zuerst wurde den Teilnehmenden jeweils für 5 Sekunden ein Bild eines Gegenstandes gezeigt. Anschließend wurde auf der linken Seite der Text „gutes Design“ und auf der rechten Seite der Text „schlechtes Design“ eingeblendet. Dieser Zustand wurde bei der *view*-Gruppe für 2 Sekunden gehalten, um sich durch Bewegung der Augen für eine Zuordnung entscheiden zu können. Nach Ablauf der 2 Sekunden wurde bei der *view*-Gruppe automatisch zu einem neuen Bild gewechselt. Die *draggable*-Gruppe hatte nach Ablauf der 5 Sekunden und dem Einblenden der Texte die Aufgabe, das Bild per Drag-and-Drop zu dem Text der ausgewählten Kategorie zu ziehen. Erst nach dem Beenden dieser Aktion wurde ein neues Bild gezeigt. Dieses Vorgehen in Phase 1 wurde für alle 120 Bilder wiederholt.

In Phase 2 wurde auf dem Screen jeweils ein Foto aus der ersten Phase und ein Objekt gleicher Kategorie aber mit anderen Eigenschaften eingeblendet (bspw. zwei unterschiedliche Scheren). Dabei wurde die Position der Objekte – rechte Seite oder linke Seite – zufällig bei jeder Darstellung eines Paares bestimmt. Durch Klicken auf eines der beiden Bilder wurde das nächste Paar eingeblendet. Dies wurde für alle 120 Paare wiederholt.

---

<sup>4</sup> DOI: 10.17605/OSF.IO/BQAH9

## 2.3 Ergebnisse

Alle Ergebnisse wurden mit Hilfe der Programmiersprache R in Version 3.6.3 berechnet und mit dem Package ggplot2 in Version 3.3.0 visualisiert. Je ein Datensatz aus jeder Gruppe wurde als Ausreißer aufgrund des >1,5fachen Interquartilsabstands nach unten zum unteren Quartil aus der Auswertung entfernt. In der Gruppe *view* konnten 19, in der Gruppe *draggable* 23 Datensätze ausgewertet werden. Die Gruppen *view* (Alter:  $M = 34.5$  Jahre,  $SD = 11.4$ ,  $Mdn = 31$ ) und *draggable* (Alter:  $M = 34.9$  Jahre,  $SD = 10.6$ ,  $Mdn = 33$ ) unterscheiden sich durch die automatische Zuordnung im Studiendesign kaum. Die geringen Unterschiede kommen durch die 2 fehlerhaften Datenerhebungen und der unter Umständen zeitlich gleichzeitigen Teilnahme an der Studie zu Stande. Es wurde die Zeit gemessen in der die Teilnehmenden die Phase 2 bewältigten. Dabei ergeben sich nur geringe Unterschiede zwischen den Gruppen: *view* (Zeit in s:  $M = 173$ ,  $SD = 31$ ,  $Mdn = 172$ ) und *draggable* (Zeit in s:  $M = 188$ ,  $SD = 44$ ,  $Mdn = 173$ ) in der Messung der Gesamtzeit. Der Median und Mittelwert liegen nahe beieinander, was zeigt, dass es keine entscheidenden Ausreißer gab.

Die Performance in visueller Gedächtnisleistung wurde durch die Anzahl der korrekt wiedererkannten Bilder der Alltagsgegenstände in Phase 2 der Studie gemessen. Jedes korrekte Zuordnen ergibt einen Punkt. Der Erwartungswert bei reiner Teilnahme an Phase 2 ohne das Durchlaufen der Phase 1 beträgt aufgrund der Konstruktion der Phase 2  $E(X) = 60$ , wobei jedes Bild mit einer Wahrscheinlichkeit von 50% gewählt wird. Die erhobene Performance ist damit aufgrund ihrer Konzeption intervallskaliert. Die adäquaten Lageparameter sind der Median und das arithmetische Mittel [FMF13], das Streumaß ist die Standardabweichung und das Zusammenhangsmaß wird im Rangkorrelationskoeffizienten angegeben, der aufgrund der kleinen Stichprobe und der großen Zahl gleicher Ränge, nach Kendall (Kendall's tau) berechnet wird [FMF13]. Nach dem Shapiro-Wilk-Test ist die Verteilung der Variablen  $Performance_{view}$ ,  $W=.77$ ,  $p>.001$  signifikant nicht normalverteilt. Bei der Verteilung der Variablen  $Performance_{draggable}$ ,  $W=.93$ ,  $p=.09$  kann von einer Normalverteilung ausgegangen werden. Aufgrund der nicht vorliegenden Normalverteilung einer Variablen wurde zur Bestimmung der Unterschiedshypothese der nichtparametrische Mann-Whitney-U-Test für unabhängige Stichproben angewendet. Für diesen Test sind alle Voraussetzungen erfüllt.

Nach den Ergebnissen des Mann-Whitney-U-Test für unabhängige Stichproben unterscheiden sich die zentralen Tendenzen der Variablen  $Performance_{draggable}$  ( $Mdn = 118$ ) und  $Performance_{view}$  ( $Mdn = 115$ ) signifikant voneinander,  $W = 317.5$ ,  $p < .05$ ,  $r = .347$ ,  $A = .727$  (vgl. Abb. 1). Die visuelle Gedächtnisleistung der Gruppe mit Touchinteraktion ( $Performance_{draggable}$ ) ist signifikant höher als diese der Gruppe ohne Touchinteraktion ( $Performance_{view}$ ). Damit wird die Nullhypothese abgelehnt und die Alternativhypothese bestätigt. Die Effektstärke ist nach Cohens Kriterium als mittlerer Effekt zu bezeichnen [FMF13].

Neben der Effektstärke bietet Vargha-Delaney  $A$  eine gute Interpretationsbasis. Vargha-Delaney  $A$  gibt ein Maß dafür an, wie oft im Mittel eine Technik der anderen überlegen



ist. Ab einem Wert von .5 dominiert eine der Techniken. Der Wert  $A = .727$  kann so interpretiert werden, dass mit einer Wahrscheinlichkeit von 73% die zusätzliche haptische Komponente in der Studie ein besseres Ergebnis in der visuellen Behaltensleistung lieferte.

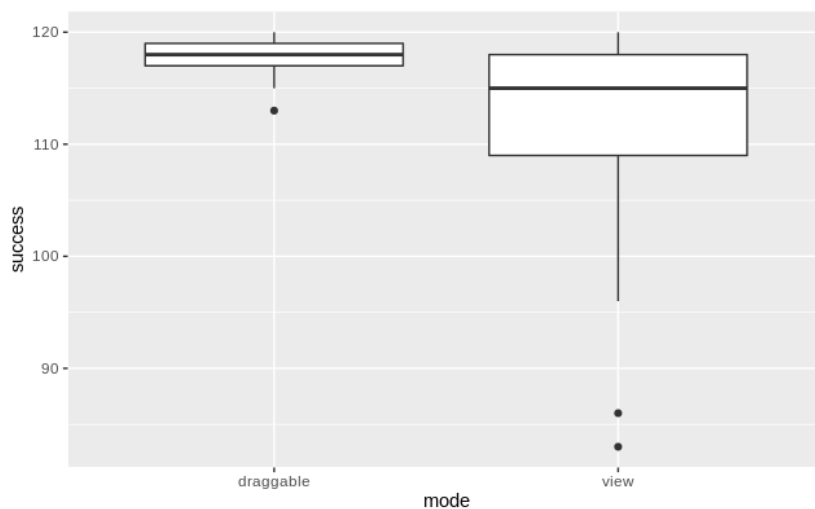


Abb. 1: Boxplot - Performance zu Gruppen *draggable* und *view*

### 3 Diskussion & Ausblick

Die Ergebnisse der vorliegenden Studie belegen, dass selbst eine minimale haptische Interaktion durch die Touch-Funktion an einem digitalen Endgerät einen Mehrwert hinsichtlich der Behaltensleistung gegenüber einer passiven Betrachtung an Standardbildschirmen bietet. Dies scheint dabei unabhängig davon zu funktionieren, ob das Memorieren des Bildschirminhalts intendiert wurde. Das Ergebnis der Studie, dass unter bestimmten Bedingungen auch eine minimale haptische Interaktion die visuelle Gedächtnisleistung verbessern kann, obwohl die Haptik nicht Bezug zu dem Objekt selbst hatte (für alle Bilder identisch war), kann als These im Bereich der Entwicklung von digitalen Lernanwendungen generalisiert und bei weiterer wissenschaftlicher Bestätigung in Design-Kriterien für digitale Lernanwendungen überführt werden. Als Schwächen der vorliegenden Studie lassen sich die geringe Kontrolle über die Teilnehmenden der Studie während der Durchführung, die fehlende Vergleichbarkeit der benutzten Endgeräte und die Gewährleistung einer adäquaten konzentrationsfördernden Umgebung nennen (aufgrund der COVID-19-Pandemie war keine Durchführung in Präsenz möglich). Aufbauend auf diesen Ergebnissen sind Studien hinsichtlich der Performance auf die visuelle Gedächtnisleistung bei Benutzung von Tangibles als haptische Eingabegeräte sowie in kollaborativen digitalen Tabletop-Szenarien geplant. Ebenso können neue Oberflächen

mit haptischer Rückmeldung sowie AR- und VR-Szenarien hinsichtlich ihrer Wirkung in dieser Forschungsfrage untersucht werden.

### Literaturverzeichnis

- [Be11] Beddall-Hill, N. et al.: Making mobile learning work. case studies of practice, 2011.
- [Br08] Brady, T. F. et al.: Visual long-term memory has a massive storage capacity for object details. In Proceedings of the National Academy of Sciences of the United States of America, 2008, 105; S. 14325–14329.
- [CH05] Castelhana, M.; Henderson, J.: Incidental visual memory for objects in scenes. In Visual Cognition, 2005, 12; S. 1017–1040.
- [Cr13] Crompton, H.: A historical overview of mobile learning: Toward learner-centered education. In (Berge, Z. L. Hrsg.): Handbook of Mobile Learning // Handbook of mobile learning. Routledge, London, 2013; S. 3–14.
- [FMF13] Field, A.; Miles, J.; Field, Z.: Discovering statistics using R. Sage, Los Angeles, Calif., 2013.
- [HK18] Hutmacher, F.; Kuhbandner, C.: Long-Term Memory for Haptically Explored Objects: Fidelity, Durability, Incidental Encoding, and Cross-Modal Transfer. In Psychological science, 2018; 956797618803644.
- [Ko10] Konkle, T. et al.: Scene memory is more detailed than you think: the role of categories in visual long-term memory. In Psychological science, 2010, 21; S. 1551–1556.
- [KRS17] Kuhbandner, C.; Rosas-Corona, E. A.; Spachholz, P.: High-Fidelity Visual Long-Term Memory within an Unattended Blink of an Eye. In Frontiers in psychology, 2017, 8; S. 1859.
- [Sa09] Salthouse, T. A.: When does age-related cognitive decline begin? In Neurobiology of Aging, 2009, 30; S. 507–514.
- [St73] Standing, L.: Learning 10,000 pictures. In The Quarterly journal of experimental psychology, 1973, 25; S. 207–222.
- [VM07] Vogt, S.; Magnussen, S.: Long-term memory for 400 pictures on a common theme. In Experimental psychology, 2007, 54; S. 298–303.
- [WHZ05] Williams, C. C.; Henderson, J. M.; Zacks, R. T.: Incidental visual memory for targets and distractors in visual search. In Perception & psychophysics, 2005, 67; S. 816–827.

## Project Heb@AR: Exploring handheld Augmented Reality training to supplement academic midwifery education

Jonas Blattgerste <sup>1</sup>, Kristina Luksch <sup>2</sup>, Carmen Lewa <sup>3</sup>, Martina Kunzendorf <sup>2</sup>, Nicola H. Bauer <sup>2</sup>, Annette Bernloehr <sup>2</sup>, Matthias Joswig <sup>3</sup>, Thorsten Schäfer <sup>3</sup>, Thies Pfeiffer <sup>1</sup>

**Abstract:** Augmented Reality (AR) promises great potential for training applications as it allows to provide the trainee with instructions and feedback that is contextualized. In recent years, AR reached a state of technical feasibility that not only allows for larger, long term evaluations, but also for explorations of its application to specific training use cases. In the BMBF funded project Heb@AR, the utilization of handheld AR as a supplementary tool for the practical training in academic midwifery education is explored. Specifically, how and where AR can be used most effectively in this context, how acceptability and accessibility for tutors and trainees can be ensured and how well emergency situations can be simulated using the technology. In this paper an overview of the Heb@AR project is provided, the goals of the project are stated and the project's research questions are discussed from a technical perspective. Furthermore, insights into the current state and the development process of the first AR training prototype are provided: The preparation of a tocolytic injection.

**Keywords:** Augmented Reality; Training; Midwifery; Education

### 1 Introduction

Augmented Reality (AR) promises great potential for training applications, as it allows to provide the trainee with instructions and feedback that is contextualized. Hereby, information is not elicited the conventional way, where a cognitive transfer process is necessary to contextualize the provided information into the current training situation, but the information is directly displayed into the physical context where and when it is needed. While it is generally well known that this can improve trainees' motivation and learning outcomes [Le12], current literature on utilizing AR for medical training specifically shows promise but often lacks the supporting evidence [BGS16]. This is at least partially explained by the limitation that AR hardware suffered from in its infancy: often technical implementations were the main challenge of projects and elaborate evaluations with technically stable prototypes have been difficult. In recent years, AR reached a state of technical feasibility that not only allows for larger, long term evaluations but also for the

---

<sup>1</sup> University of Applied Sciences Emden/Leer (HSEL), Constantinaplatz 4, 26721 Emden, Germany,  
vorname.nachname@hs-emden-leer.de

<sup>2</sup> Hochschule für Gesundheit (hsg Bochum), Gesundheitscampus 6-8, 44801 Bochum, Germany,  
vorname.nachname@hs-gesundheit.de

<sup>3</sup> Ruhr-Universität Bochum (RUB), Universitätsstraße 150, 44801 Bochum, Germany,  
vorname.nachname@ruhr-uni-bochum.de

exploration of the technology in those specific training use cases, e.g. in the medical context. Now the focus of projects shift from the technical feasibility to identifying and exploring needs of specific contexts, evaluating developed prototypes with larger sample sizes and starting the process of exploring long term implementations and acceptances of AR applications in existing training structures. Increasingly, AR based training can be explored as an extension of computer- & web-based trainings, that are already successfully used for decades, to extend them with the utility of contextualizing information that can be provided by AR based training applications (see Fig. 1).

In the BMBF funded project Heb@AR, the utilization of handheld AR as a supplementary tool for the practical training in the academic midwifery education is explored, which is not yet covered by the literature. Specifically of interest is how and where can AR be used effectively in this context, how can acceptability and accessibility for tutors and trainees be ensured and how well can emergency situations be simulated using the technology.



Fig. 1: AR-based training is the logical extension of computer-based & web-based training

## 2 Project Overview

The German midwifery education is currently transitioning towards a full academization. Based on a recent law, from 2022 onwards, midwives will be exclusively qualified at universities, rather than by vocational training through the dual education system, which is still the standard today [HG20]. While this is an important step towards increasing the status of midwives in the medical context, it also leads to new challenges. The practical component of the training still has a high priority with exemplary bachelor's consisting of 4380 hours of theoretical and 2200 hours of practical training. This naturally leads to bottlenecks regarding available practical tutors, training space and scheduling restrictions for trainees in the long run. Furthermore, with the full academization of the midwifery education, the heterogeneity of students magnifies, only increasing the impact of those restrictions due to potential needs for more individual support. Despite those new challenges, the general goals remain: Students have to reach a theoretical and technical expertise and especially the transfer from theory to practice has to succeed. Furthermore,

students also have to be prepared for important emergency situations that cannot be trained reliably in practice. In general, the academic education is moving towards problem-based learning approaches to enable students more self-regulated learning, often in collaborative settings [Zu12]. This is partially done in laboratory training sessions and with the utilization of learning-management systems. The development of problem-based learning approaches for midwifery students in particular hereby faces special challenges and is largely unexplored but could have huge potential based on the high percentage of practical training.

Starting in November 2019, forces have joined with interdisciplinary researchers from midwifery science, medical-didactics and computer science for a duration of three years to implement and evaluate AR-based training as a supplementary tool for problem-based learning approaches in midwifery education. In short, the project will consist of the development of 3 exemplary AR training scenarios from the field of emergency management for handheld AR devices that are iteratively evaluated and improved, the implementation of those AR scenarios into existing structures (e.g. learning-management systems) and ultimately the exploration of methods to enable tutors themselves to create new AR trainings. Within the project several interesting research questions arise for each project partner. The ones with a technical primary focus are highlighted in the following.

1. **Acceptability:** How can handheld AR be utilized effectively and implemented in a way that it is intuitively usable and perceived as useful by the students?
2. **Scalability:** Which scenarios are suitable to be trained through handheld AR? Do AR scenarios scale better with the increasing number of students compared to on-site trainings? Could AR even be used to realize novel exam formats?
3. **Viability & Longevity:** How can tutors/lecturers be enabled to create their own scenarios using authoring tools, which was identified as a crucial factor for successful longer-term AR implementations in previous work [BRP19]? How can handheld AR be integrated seamlessly into existing technical (e.g. learning-management systems) but also social teaching and training structures and contexts?
4. **Self-regulation & Collaboration:** In which ways can handheld AR support self-regulated and collaborative learning at home and in training facilities?

Beyond that, the researchers from midwifery science are especially, but not comprehensively, interested in the transferability of acquired knowledge and the aspect of interdisciplinary cooperation. Therefore, how can knowledge and skills acquired in AR based training be applied in practice? How effective can this transfer of knowledge be elicited? Can AR improve interdisciplinary cooperation in obstetric emergencies? In obstetric emergencies, at least two professional groups are usually involved in the care. Interprofessional communication and action are important. The medical-didactic researchers additionally interest the identification of factors regarding an interprofessional and interdisciplinary transferability of the AR formats.

### 3 Development of the first prototype

In Summer 2020, the first prototype for one of the three scenarios is developed: The preparation of a tocolytic injection, which is currently being evaluated. Here, a syringe with a carrier solution and a drug is prepared and stored for emergency situations.

While the project follows a design-based research approach in accordance with [RG05], the procedure that was used for bootstrapping the first prototype differs slightly, as additional challenges emerged due to the exploratory nature: It had to be ensured that the action sequence that will be implemented is correct and that didactic considerations are integrated while still retaining good overall usability of the application. Therefore, a more linear approach for this exploration was chosen. Further improvements to the prototype will be made iteratively based on feedback.

First, the action sequence that was supposed to be implemented was observed in detail and each step analyzed to ensure correctness. The action sequence was then furthermore recorded on video and shared with the didactic and technical implementation team. Based on the observed and recorded action sequence, a task process analysis [JHT89] was created, formalizing the action sequence.

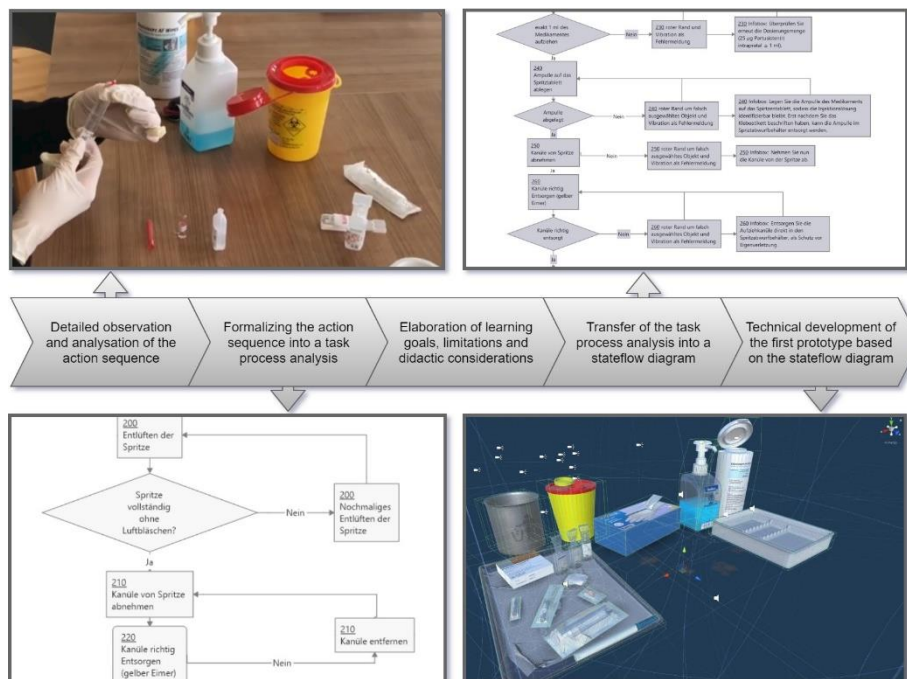


Fig. 2: The development process of our first AR prototype for the AR training scenario of preparing a tocolytic injection.

Based on the task process analysis, the learning goals, didactic considerations and technical limitations were discussed and formalized. Utilizing this, the process analysis was translated into a stateflow diagram, specifying the technical implementation more closely and was ultimately implemented to be the first prototype. Our development process is visualized with some exemplary visualizations in Fig. 2.

Simultaneously, during this process, possible interaction concepts mitigating the limitations for handheld AR (e.g. having one hand occupied) and the required 3D models were explored and prototyped on paper, with low-poly models and then realized as high-poly versions with complex interactions, physics, shadows and transparency effects (see Fig. 3).

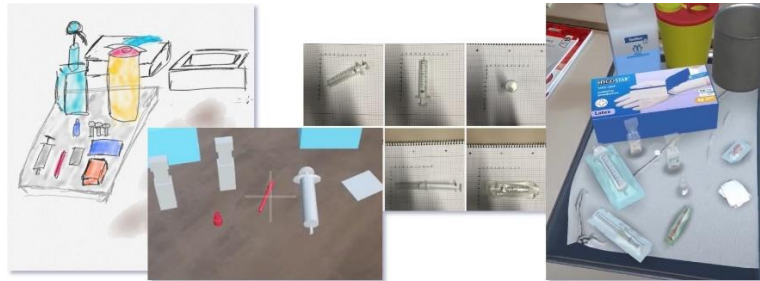


Fig. 3: Examples of the prototyping process of models and their interactions from paper prototyping, through low-poly models with simple interactions, to modeling the final versions.

The current state of the prototype is visualized in Fig. 4. It allows for location-independent preparation of the injection using only virtual objects that can be grabbed, combined and interacted with through movement of the smartphone and on-screen touch interactions. This first scenario in the project is also the onboarding scenario for our experimental group and is therefore relatively simple in its objective compared to the following scenarios.



Fig. 4: The AR prototype guiding the virtual preparing of a tocolytic injection. Displayed is the step of drawing up the syringe with a carrier solution.

## 4 Conclusion & Outlook

In this paper the BMBF funded project Heb@AR was described, where the usage of handheld AR training to supplement the practical phases of academic midwifery education is explored. The project itself, the projects goals and the research questions were described on an abstract level. Furthermore, the current state and development process of the first prototype, the preparation of a tocolytic injection, was described.

In the future, it is planned to iteratively evaluate and improve the existing and upcoming prototypes for the AR trainings and to develop an authoring tool, enabling tutors to create their own AR trainings. Furthermore, the usability and reception of the interaction concept that was developed for the implementation of the procedural training tasks on handheld AR devices, which is only incidentally described in this paper, will be evaluated and published separately. In an ongoing usability study, the interaction concept and required extend and form of interaction onboarding is explored through a task-based, comparative, observational usability study with midwifery students that already completed the module the AR training is targeting to supplement. Here, the independent variables are the form and extend of interaction onboarding that is provided before using the AR training, which inherently also provides first insights into the usability of the interaction concept itself.

## Bibliography

- [HG20] Gesetz über das Studium und den Beruf von Hebammen (Hebammengesetz - HebG) (2020) in der Fassung des Gesetzes zur Reform der Hebammenausbildung und zur Änderung des Fünften Buches Sozialgesetzbuch (Hebammenreformgesetzes- HebRefG) vom 22.11.2019 (BGBl. I S. 1759), 2020.
- [BGS16] Barsom, E. Z.; Graafland, M.; Schijven, M. P.: Systematic review on the effectiveness of augmented reality applications in medical training. *Surgical endoscopy* 30/10, S. 4174–4183, 2016.
- [RG05] Reinmann, Gabi. "Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung." *Unterrichtswissenschaft* 33.1, 2005.
- [BRP19] Blattgerste, J.; Renner, P.; Pfeiffer, T.: Authorable augmented reality instructions for assistance and training in work environments. In: *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. S. 1–11, 2019.
- [JHT89] Jonassen, D. H.; Hannum, W. H.; Tessmer, M.: *Handbook of task analysis procedures*. Praeger Publishers, 1989.
- [Le12] Lee, K.: Augmented reality in education and training. *TechTrends* 56/2, S. 13–21, 2012.
- [Zu12] Zumbach, J. et al.: *Problem-Based Learning in Österreich: eine Bestandsaufnahme 2012*



## Aufgaben mit automatischem Feedback zu chemischen Atom-Orbitalschemata

Michael Striewe<sup>1</sup>, Florian Trauten<sup>2</sup> und Carolin Eitemüller<sup>3</sup>

**Abstract:** Übungs- und Prüfungsaufgaben enthalten oft fachspezifische Notationen oder Interaktionsformen, die bei einer Umsetzung in elektronische Systeme beachtet werden müssen. Nicht immer stehen dabei passende Komponenten für eine exakte Übertragung zur Verfügung und auch eine Umsetzung mit abgewandelten Aufgabentypen ist nicht immer sinnvoll möglich. An einem Beispiel aus der allgemeinen Chemie diskutiert der Beitrag, wie eine fachspezifische Eingabemöglichkeit in einem gegebenen E-Assessment-System geschaffen werden kann und welche didaktischen Überlegungen zur Benutzbarkeit und automatischen Feedbackerzeugung dabei beachtet werden mussten. Eine Evaluation im Einsatz belegt, dass in beiden Punkten ein für die Lernenden ausreichendes Ergebnis erzielt werden konnte.

**Keywords:** E-Assessment, Fachspezifische Eingaben, Automatisches Feedback, Chemie

### 1 Einleitung und Motivation

Zu den ständigen Herausforderungen bei der Übertragung vorhandener, analoger Lernmaterialien in technische Systeme gehört es, fachspezifische Notationen oder Interaktionsformen auf geeignete Weise in das eingesetzte, digitale Lernwerkzeug zu übertragen. In einfachen Fällen kann dazu eine Umsetzung mit standardisierten Aufgabentypen wie Multiple-Choice oder Lückentexten möglich sein, ohne dabei Einbußen in der Qualität der Aufgabe oder des möglichen Feedbacks hinzunehmen. Selbst starke Änderungen wie das Ersetzen einer Zeichenaufgabe durch das Auswählen einer vorgegebenen Grafik wirken sich nicht zwingend auf das Ergebnis aus [ABKY05], ändern aber gleichwohl den Charakter der Aufgabe. Ein weiterer Ansatz besteht darin, generische Eingaben für fachspezifische Artefakte zu verwenden, beispielsweise chemische Formeln in Form von eingereichten Grafiken entgegenzunehmen [PBR07]. In diesem Fall wird jedoch eine automatische Bewertung zumindest erschwert, wenn nicht sogar verhindert. Wenn sowohl die spürbare Veränderung der Aufgabe durch die Wahl eines anderen Aufgabentyps als auch der Verzicht auf die automatische Bewertung durch ein digitales Lernwerkzeug nicht akzeptabel ist, bleibt als weitere Option nur die Entwicklung zusätzlicher Features für das verwendete digitale Lernwerkzeug.

---

<sup>1</sup> Universität Duisburg-Essen, paluno – The Ruhr Institute for Software Technology, Gerlingstraße 16, 45127 Essen, michael.striewe@paluno.uni-due.de

<sup>2</sup> Universität Duisburg-Essen, Didaktik der Chemie, Schützenbahn 70, 45127 Essen, florian.trauten@uni-due.de

<sup>3</sup> Universität Duisburg-Essen, Didaktik der Chemie, Schützenbahn 70, 45127 Essen, carolin.eitemueller@uni-due.de

In der Chemie wurde dieser Weg ebenfalls schon mehrfach und mit unterschiedlichem Erfolg besritten: Verschiedene Möglichkeiten der Eingabe chemischer Strukturformeln haben sich bisher eher als unzureichend erwiesen [Jo15], während eine Erweiterung eines mathematischen Formeleditors um Elemente und Auswertungsmöglichkeiten für chemische Reaktionsgleichungen erfolgreicher verlief [PS19]. In beiden Fällen konnten als Arbeitsgrundlage vorhandene Software-Lösungen erprobt und (im zweiten Fall) erfolgreich erweitert werden. Je spezialisierter das Thema einer Aufgabe ist, umso unwahrscheinlich ist es jedoch, dass eine solche Wiederverwendung möglich ist.

Der vorliegende Beitrag befasst sich mit genau einem solchen Fall und geht der Frage nach, ob auch ein Aufgabentyp zu chemischen Atom-Orbitalschemata und deren grafischer Repräsentation digital umgesetzt werden kann. Im Zentrum stehen dabei zwei Fragen: Kann für diesen speziellen Aufgabentyp ein web-basierter Editor entwickelt werden, der für Studierende hinreichend intuitiv bedienbar ist und kann die Auswertung der Antworten soweit automatisiert werden, dass ein nach didaktischen Gesichtspunkten gestaltetes, differenziertes Feedback möglich ist?

## 2 Atom-Orbitalschemata

Die in diesem Beitrag betrachteten Atom-Orbitalschemata sind eine Diagrammform, in der verschiedene Informationen über die Verteilung der Elektronen in einem Atom kodiert werden. Das Atom-Orbitalschema lässt sich auf der Schrödinger-Gleichung begründen. Jedem energetischen Zustand, den ein Elektron einnehmen kann, lässt sich eine definierte Wellenfunktion  $\Psi$  zuordnen. „Die Wellenfunktion für ein Elektron in einem Atom ist der mathematische Ausdruck für etwas, das wir [Atom-]Orbital nennen“ [MM14]. Mithilfe von drei Quantenzahlen lassen sich bei der wellenmechanischen Betrachtung des Elektrons die Aufenthaltsbereiche und die begrenzenden Knotenflächen dieser Bereiche charakterisieren.

Die Hauptquantenzahl  $n$  (mit  $n = 0, 1, 2, 3, \dots$ ) definiert die sog. Schale (K, L, M, ...; synonym zur Periode im Periodensystem), in der sich ein Elektron mit großer Wahrscheinlichkeit aufhalten wird. Dabei hängt die Größe des Zahlenwertes proportional mit der Entfernung des Elektrons zum Kern zusammen. Jede Schale kann wiederum über die Nebenquantenzahl  $l$  (mit  $l = 0, 1, 2, \dots (n - 1)$ ) in Unterschalen aufgeteilt werden. Diese benennen den Orbitaltyp (s, p, d, ...). Jeder Nebenquantenzahl wird zudem noch eine Magnetquantenzahl  $m$  bzw.  $l_m$  (mit  $l_m = -l, -(l - 1), \dots, 0, \dots (l + 1), l$ ) zugeordnet. Diese Zahl gibt an, wie viele Orbitale es in dieser Unterschale gibt. S-Orbitale haben mit  $l=0$  somit auch eine Magnetquantenzahl von  $l_m=0$ , wodurch sich ergibt, dass es in jeder Schale nur ein s-Orbital gibt. P-Orbitale haben die Nebenquantenzahl  $l=1$  und somit  $l_m=\{-1,0,1\}$ , so dass es stets drei p-Orbitale gibt, wobei diese erst ab der L-Schale auftreten. Zuletzt gilt es zu erwähnen, dass jedes Orbital mit zwei Elektronen besetzt werden kann, welche jedoch mit entgegengesetztem Spin (einem Eigendrehimpuls) ein magnetisches Moment erzeugen. Die sogenannte

Spinquantenzahl  $s$  beschreibt also nicht die Form eines Orbitals, sondern nur das magnetische Moment eines Elektrons, wobei zwei Zustände möglich sind  $\left(+\frac{1}{2}, -\frac{1}{2}\right)$ .

Die Belegung der Orbitale ist dabei zwei Gesetzmäßigkeiten unterworfen: Nach dem Pauli-Prinzip dürfen keine zwei Elektronen eines Atoms in allen vier Quantenzahlen übereinstimmen. Das heißt innerhalb desselben Orbitals derselben Schale darf es keine zwei Elektronen mit demselben Spin geben. Nach der Hund'schen Regel verteilen sich Elektronen bei Orbitalen gleichen Energieniveaus zudem so, dass ein Maximum ungepaarter Elektronen mit parallelem Spin entsteht. Es darf also nicht gleichzeitig in einer Unterschale leere Orbitale und Orbitale mit zwei Elektronen geben. Die Kenntnis beider Gesetzmäßigkeiten und der sich daraus ergebenden Konsequenzen für den Aufbau eines Atoms sind ein wichtiges Lernziel in der Einführung in die allgemeine Chemie.

Die Verteilung der Elektronen auf die Schalen und die darin befindlichen Orbitale lässt sich schematisch in einem sogenannten Atom-Orbitalschema darstellen. Der Fokus liegt bei dieser abstrakten Visualisierungsform auf der Verteilung der Elektronen, die räumliche Ausrichtung der Orbitale bleibt hier unbehandelt. Die Position der Orbitale entlang der y-Achse gibt dabei Auskunft über das (relative) Energieniveau. Abbildung 1 zeigt ein solches Atom-Orbitalschema für ein einzelnes Sauerstoffatom. Dem Diagramm ist zu entnehmen, dass das Sauerstoffatom über acht Elektronen verfügt, von denen sich zwei im sogenannten 1s-Orbital in der K-Schale, sowie zwei weitere in der L-Schale im 2s-Orbital befinden und die übrigen sich möglichst gleichmäßig auf die drei verfügbaren 2p-Orbitale verteilen. Dabei hat das 2p-Orbital ein höheres Energieniveau als das 2s-Orbital und dieses wiederum ein höheres Energieniveau als das 1s-Orbital. Anhand der Pfeile kann außerdem noch die Spin-Richtung der Elektronen abgelesen werden.

Alle diese Informationen sind in verschiedenen Kontexten relevant, um beispielsweise das Verhalten verschiedener Elemente in chemischen Reaktionen erklären zu können. (Atom-)Orbitalschemata sind daher ein wichtiger Baustein, um Faktenwissen über einzelne Elemente in unmittelbarer Verbindung mit der Theorie zum Aufbau von Atomen zu vermitteln, darzustellen und zu prüfen. Sowohl das Faktenwissen alleine als auch der theoretische Aufbau von Atomen kann auch separat mit rein textuellen Darstellungen vermittelt bzw. geprüft werden, nicht jedoch die praktische Verbindung der beiden Aspekte. Selbst wenn man dazu rein textuelle Aufgaben stellen würde, erfordert eine Beantwortung im Wesentlichen, zunächst ein Orbitalschema zu erstellen und dann die gewünschten Informationen abzuleiten.

Praktisch ergeben sich dabei mindestens zwei Aufgabentypen, bei denen Atom-Orbitalschemata zum Einsatz kommen können: Erstens die Erstellung von Atom-Orbitalschemata für ein gegebenes Atom und zweitens die Ableitung von Informationen aus einem gegebenen Atom-Orbitalschema. Der vorliegende Beitrag betrachtet nur den ersten Aufgabentyp, da im zweiten Fall das Atom-Orbitalschema durch Lehrende bzw. Aufgabenautoren erstellt wird und sich der Aufgabentyp somit nicht wesentlich von beliebigen anderen Aufgaben unterscheidet, in denen Lernende eine Abbildung vorgelegt bekommen.

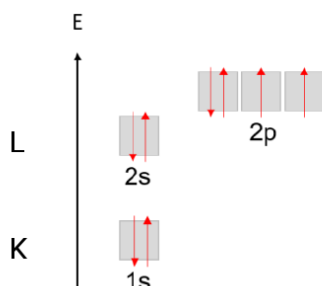


Abb. 1: Atom-Orbitalschema für Sauerstoff.

### 3 Entwicklung des Editors

Rahmenbedingung für die Entwicklung des Editors war die Einbindung in das E-Assessment-System JACK, welches an der Universität Duisburg-Essen als eines von zwei dedizierten E-Assessment-Systemen (zusätzlich zu einem allgemeinen Lern-Management-System) regelmäßig und fakultätsübergreifend im Einsatz ist. Insbesondere werden auch andere Übungsaufgaben in der allgemeinen Chemie über dieses System abgewickelt, so dass die Studierenden im allgemeinen Umgang mit dem System vertraut sind. Das System ist web-basiert und der zu entwickelnde Editor sollte in möglichst vielen modernen Browsern und auf möglichst vielen Geräten nutzbar sein. Daraus ergab sich schnell die Entscheidung für eine Umsetzung mit HTML5 und JavaScript, so dass der Editor grundsätzlich auch in andere Systeme übertragen werden kann. Die Überprüfung der Lösung und die Erzeugung von Feedback ist serverseitig in Java realisiert und damit weniger leicht auf andere Systeme übertragbar.

#### 3.1 Anforderungen an Funktion und Usability

Der benötigte Funktionsumfang für den Editor ist verhältnismäßig gering. Es muss für den Benutzer möglich sein, Schalen und Orbitale hinzuzufügen und wieder zu entfernen. Ferner muss es möglich sein, Elektronen in vorhandenen Orbitalen zu platzieren und dabei ihren Spin auszuwählen. Außerdem müssen Elektronen auch wieder entfernt werden können.

Aus didaktischen Gründen wurden zudem einige Vereinfachungen vorgenommen, so dass weitere, naheliegende Funktionen zunächst nicht realisiert werden mussten. Dies betrifft insbesondere die Positionierung der Orbitale, so dass es den Lernenden weder möglich sein sollte, Orbitale und Schalen in einer falschen Reihenfolge zu positionieren, noch ihnen ein falsches Energieniveau zuzuweisen. Es soll jedoch möglich sein, real nichtexistente Orbitale hinzuzufügen (z. B. d-Orbitale in der K-Schale). Dies bedingt auch, dass diesen Orbitalen dann automatisch eine schlüssige Position im Diagramm

zugewiesen werden muss. Eine vereinfachte Darstellung, die für jede Schale ein einheitliches Energieniveau annimmt, wurde dabei zunächst als ausreichend angesehen.

Um auf möglichst vielen Browsern und Geräten gleichförmig bedienbar zu sein, sollte der Editor möglichst wenig verschiedene Bedienaktionen erfordern. Im Idealfall sollte die Bedienung ausschließlich per „Point-and-Click“ möglich sein, so dass weder mehrere Maustasten noch komplexe Mausbewegungen bzw. Fingergesten und auch keine Tastatureingaben nötig sind. Dies wiederum impliziert einige Vorgaben, indem die Benennung von Schalen (K, L, M, ...) und Orbitalen (s, p, d, ...) in Auswahlfeldern zur Verfügung steht und somit eine völlig falsche Benennung grundsätzlich ausgeschlossen ist.

### **3.2 Anforderungen an die Feedbackgenerierung**

Um die Korrektheit eines Atom-Orbitalschemas zu prüfen, muss lediglich die Ordnungszahl des dargestellten Elements bekannt sein und es müssen einige chemische Gesetze beachtet werden, insbesondere das oben bereits erwähnte Pauli-Prinzip und die Hund'sche Regel. Es ist daher nicht notwendig, dass Lehrende bzw. Aufgabenautoren individuelle Prüfregeln für verschiedene Elemente festlegen können, sondern die Überprüfung soll vollständig automatisch durchgeführt werden können. Zu beachten ist dabei jedoch, dass es im Allgemeinen mehrere korrekte Lösungen einer Aufgabe geben kann. Dazu ist es lediglich notwendig, in einem gegebenen, korrekten Atom-Orbitalschema alle Spins aller Elektronen umzukehren, um so ein zweites, korrektes Diagramm zu erhalten. Das Umkehren des Spins in einer Teilmenge der Elektronen kann dagegen auch zu falschen Lösungen führen. In bestimmten Konstellationen ist es auch möglich, einzelne Elektronen an verschiedenen Stellen zu platzieren, die alle korrekt sind. Bei der Feedbackerzeugung müssen daher tatsächlich verschiedene Regeln überprüft werden und es kann nicht einfach mit einer Musterlösung verglichen werden. Zudem soll detailliertes Feedback erzeugt werden, das den vorhandenen Fehler möglichst genau benennt, beispielsweise durch die Benennung der verletzten chemischen Gesetzmäßigkeit. Bei der Berechnung einer Punktzahl für eine Lösung sollen verschiedene Fehler zudem unterschiedlich gewichtet werden können.

### **3.3 Initialer Prototyp**

Ein erster Prototyp eines geeigneten Editors wurde Anfang 2018 entwickelt und in den darauffolgenden Semestern evaluiert. Abbildung 2 zeigt einen Screenshot der Oberfläche. Diese besteht aus dem grau hinterlegten Diagrammbereich und einem darunterliegenden Bereich mit Steuerelementen. Zur Bearbeitung des Diagramms muss ausschließlich dieser untere Bereich bedient werden. Buttons und Drop-Down-Menüs können dabei wie geplant ausschließlich mit der Maus bedient werden. In zwei Eingabefeldern sind jedoch Tastatureingaben notwendig. Der Prototyp erfüllt damit die funktionalen Anforderungen vollständig, die Anforderungen an die Usability jedoch nicht gänzlich.

Technisch manipuliert die Interaktion mit dem Bedienfeld eine Objektstruktur, die das Atom-Orbitalschema repräsentiert. Diese Objektstruktur wird dann einerseits in der Diagrammfläche grafisch umgesetzt und andererseits in Form eines JSON-Strings serialisiert und in ein verstecktes Eingabefeld der Webseite geschrieben, um auf diesem Weg bei einer Einreichung zum Server geschickt zu werden. Für die automatische Feedbackerzeugung wird die Datenstruktur dann dort analysiert.

**Aufgabenbeschreibung:** Modellieren sie die Elektronenkonfiguration des Sauerstoffatoms.

Abb. 2: Benutzeroberfläche der prototypischen Implementierung des Atom-Orbitalschema-Editors mit unvollständig eingetragenen Sauerstoff-Atom. Die Energieniveaus werden in dieser Version nur vereinfacht dargestellt.

### 3.4 Evaluation des Prototyps und weitere Anforderungen

Zur Evaluation des Prototyps wurden Aufgaben zu verschiedenen Atomen erstellt und in Sachen Bedienbarkeit und Vollständigkeit des Feedbacks von zwei Experten der Chemiedidaktik evaluiert. Die Bedienbarkeit wurde dabei als hinreichend gut eingestuft, so dass sich nur kleine Änderungsanforderungen zum Layout ergaben. Weitere Ideen, z. B. zur Auswahl von Orbitalen in der Diagrammfläche durch Mausklick wurden zwar diskutiert, aber als nicht dringend eingestuft.

In der Diagrammfläche ergab sich die Beobachtung, dass der Editor eine zu stark vereinfachte Darstellung wählt, indem er alle Orbitale einer Schale auf demselben Energieniveau darstellt. Diese Art der Darstellung war zwar in den initialen Anforderungen als akzeptabel angesehen worden, erwies sich in der Evaluation aber

doch als störend. Insbesondere mussten zur Erzeugung des Feedbacks mehr Details berücksichtigt werden, als für die Studierenden im Diagramm ersichtlich waren. Daraus ergab sich als neue Anforderung, dass für alle Orbitale ein korrektes, relatives Energieniveau bei der Anzeige berücksichtigt werden muss. Das betrifft - wie oben bereits diskutiert - auch fälschlich eingefügte, real nicht existierende Orbitale, die trotzdem eine plausible Position erhalten müssen. Eine freie Positionierung der Orbitale durch die Studierenden wurde in Bezug auf die Bedienbarkeit als unverhältnismäßig komplex eingeschätzt und daher abgelehnt. Ferner ergaben sich kleinere Änderungsanforderungen zu Inhalt und Position der Labels zur Benennung der Orbitale.

In Bezug auf die Feedbackerzeugung wurden einige Fälle entdeckt, für die noch kein ausreichendes Feedback erzeugt wurde. Gleichzeitig stellte sich heraus, dass eine Gewichtung jeder einzelnen Regel bei der Berechnung einer Punktzahl nicht notwendig ist. Stattdessen erschien es ausreichend, zwischen falsch gewählten Bestandteilen (z. B. zu viele oder zu wenige Elektronen, nicht existierende Orbitale, usw.) und falsch angewandten Gesetzen (z. B. falsche Verteilung der Elektronen auf die Orbitale, falscher Spin, usw.) zu unterscheiden. Ferner sollte es zusätzlich möglich sein, dass Lehrende das automatische, generische Feedback um aufgabenspezifische Texte ergänzen, um beispielsweise individuell auf Lehrmaterial oder andere Beispiele verweisen zu können.

### 3.5 Weiterentwicklung

Die Umsetzung der neuen Anforderungen wurde Mitte 2019 durchgeführt. Abbildung 3 zeigt die nur leicht veränderte Benutzeroberfläche. Keine der neuen oder geänderten Anforderungen erforderte eine umfassende Änderung der Datenstrukturen oder der Implementierung. Erweiterungen waren jedoch sowohl client-seitig notwendig, um das Energieniveau zu berücksichtigen, als auch serverseitig, um das erweiterte Feedback zu realisieren.

## 4 Evaluation im Einsatz

Im Wintersemester 2019/20 wurde der Editor erstmals mit Studierenden im regulären Übungsbetrieb einer Einführung in die allgemeine Chemie eingesetzt. Insgesamt wurden sechs Aufgaben zur freiwilligen (ggf. auch mehrfachen) Bearbeitung angeboten. Im Rahmen einer umfangreicheren Studie sollte insbesondere der Effekt von automatisch generiertem, detailliertem Feedback für verschiedene Aufgabentypen untersucht werden. In allen Aufgabentypen war dabei eine fachspezifische Eingabe notwendig (insbesondere auch die in der Einleitung erwähnte Eingabe von chemischen Reaktionsgleichungen), so dass auch die Einschätzung der Studierenden zur Nutzbarkeit der Eingabemöglichkeiten sowie ihre empfundene kognitive Belastung bei der Bearbeitung der Aufgabe erhoben wurden, um Vergleiche ziehen zu können. Der vorliegende Beitrag befasst sich jedoch ausschließlich mit den Ergebnissen zum Aufgabentyp der Atom-Orbitalschemata.

Geben Sie die Orbitale und Elektronenbelegungen von Sauerstoff an:

Orbital hinzufügen	Orbital entfernen	Elektronenanzahl ändern	Allgemein
Schale: <input type="text" value="L"/>	Schale: <input type="text" value="K"/>	Schale: <input type="text" value="L"/>	<input type="button" value="Reset"/>
Orbitalbezeichnung: <input type="text" value="p"/>	Orbitalbezeichnung: <input type="text" value="s"/>	Orbital(e): <input type="text" value="p"/>	
Orbitalanzahl: <input type="text" value="3"/>		Orbitalnummer: <input type="text" value="1"/>	
<input type="button" value="+"/>	<input type="button" value="-"/>	<input type="button" value="↑"/> <input type="button" value="↓"/> <input type="button" value="entf"/>	

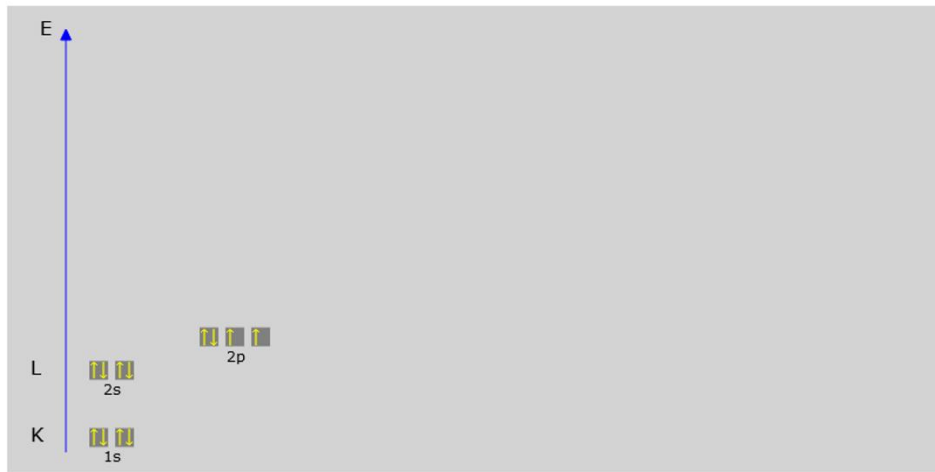
Abb. 3: Benutzeroberfläche der finalen Implementierung des Atom-Orbitalschema-Editors mit unvollständig eingetragem Sauerstoff-Atom.

Zum Zweck der Studie wurden die Studierenden zufällig in zwei Gruppen eingeteilt. Die erste Gruppe, aus der 25 Studierende insgesamt 85 Bearbeitungen der sechs angebotenen Aufgaben erzeugten, erhielt dabei detailliertes, automatisches Feedback. Die zweite Gruppe, aus der 14 Studierende insgesamt 36 Bearbeitungen erzeugten, wurde nur über die Korrektheit der Antwort informiert. In beiden Gruppen erhielten die Studierenden in jeder Aufgabe dieselben Hinweise zur Bedienung des Atom-Orbitalschema-Editors. Ferner wurde allen Studierenden nach drei Fehlversuchen in einer Aufgabe eine Musterlösung angezeigt.

Ein Beispiel für einen falschen Lösungsversuch und das dazu automatisch erzeugte Feedback ist in Abbildung 4 zu sehen. Die Lösung weist zwei, voneinander unabhängige Fehler auf: Zum einen sind jeweils zwei statt einem 1s- bzw. 2s-Orbital eingezeichnet. Zum anderen sind insgesamt 12 statt 8 Elektronen eingetragen worden. Zu beiden Fehlern wurde ein entsprechendes Feedback erzeugt.

Alle Studierenden wurden per Fragebogen hinsichtlich der Aspekte Usability, empfundene Aufgabenschwierigkeit und Denkanstrengung zum Lösen der Aufgabe befragt. Die Studierenden, die detailliertes Feedback erhalten hatten, wurden zudem zur Nützlichkeit des Feedbacks befragt. Insgesamt nahmen 31 Studierende an der freiwilligen Befragung teil, davon 19 aus der Gruppe mit detailliertem Feedback. Die Ergebnisse der Befragung sind in Tabelle 1 und 2 aufgeführt.



**Feedback:****Die Antwort ist leider falsch.**

Im Diagramm sind mehr Elektronen dargestellt, als dem Sauerstoffatom zur Verfügung stehen. Um die korrekte Anzahl an Elektronen zu ermitteln betrachten Sie die Ordnungszahl des Elements und leiten Sie daraus die Anzahl an Elektronen ab.

Im Diagramm sind Orbitale dargestellt, die in der von Ihnen gewählten Schale nicht existieren. Um die korrekte Anzahl an Elektronen zu ermitteln betrachten Sie die Ordnungszahl des Elements und leiten Sie daraus die Anzahl an Elektronen ab. Füllen Sie dann mithilfe der Visualisierung der Hund'schen Regel und unter Berücksichtigung des Pauli-Prinzips die Orbitale mit Elektronen.

Abb. 4: Beispiel einer falschen Lösung und des dazu automatisch erzeugten Feedbacks.

Alle Studierenden schätzte die Aufgabe als „sehr leicht“ oder zumindest „leicht“ ein. Passend dazu empfanden 28 Studierende (90%) die Denkanstrengung als „gering“ oder sogar „sehr gering“. Daraus kann geschlossen werden, dass der Aufgabentyp als solcher weder grundsätzlich noch in der spezifischen, technischen Aufbereitung eine nennenswerte Herausforderung für die Studierenden darstellt. Dabei zeigten sich keine Unterschiede zwischen der Gruppe mit und ohne detailliertem Feedback.

Der Umgang mit der Bedienoberfläche der Aufgabe wurde von 20 Studierenden (65%) positiv als „eher intuitiv“ bis „sehr intuitiv“ bewertet. Von den Studierenden, die automatisiertes fehlerspezifisches Feedback bei der Aufgabe erhalten hat, empfanden 13 Studierende (68 %) das Feedback als „eher hilfreich“ bis „sehr hilfreich“. Damit kann für die technischen Aspekte eine zumindest ausreichend gute Lösung angenommen werden, die allerdings gleichwohl noch Potenzial für Verbesserungen offenbart. Zur Ableitung konkreter Verbesserungen sind die Ergebnisse jedoch nicht detailliert genug. Dazu sollte stattdessen eine dedizierte Usability-Studie durchgeführt werden, in der die Schwierigkeiten bei der Interaktion mit dem Editor genauer erhoben werden.

	sehr leicht	leicht	schwierig	sehr schwierig
Wie schwierig waren die Aufgaben, die Sie gerade bearbeitet haben, zu lösen?	16	15	0	0
	sehr gering	gering	hoch	sehr hoch
Bei der Bearbeitung der Aufgaben war meine geistige Denkanstrengung insgesamt ...	12	16	3	0
	sehr intuitiv	eher intuitiv	eher nicht intuitiv	gar nicht intuitiv
Wie intuitiv haben Sie die Eingabe bei den Aufgaben gefunden?	5	15	6	5

Tab. 1: Ergebnisse der Befragung aller Studierenden

	sehr hilfreich	eher hilfreich	eher nicht hilfreich	gar nicht hilfreich
Wie hilfreich war das Feedback bei der Lösung der Aufgaben?	5	8	4	2

Tab. 2: Ergebnisse der Befragung der Studierenden, die detailliertes Feedback erhalten haben.

## 5 Fazit und Ausblick

In diesem Beitrag konnte gezeigt werden, dass mit verhältnismäßig geringem Aufwand eine technische Lösung für die fachspezifische Eingabe und automatische Feedback-erzeugung entwickelt werden konnte, die die didaktischen Anforderungen berücksichtigt und die von den Lernenden zumindest als ausreichend bewertet wird. Damit konnte ein Aufgabentyp zu Atom-Orbitalschemata in einem E-Assessment-System umgesetzt werden, ohne dazu eine vereinfachte Interaktionsform (z. B. Multiple-Choice) verwenden oder auf die automatische Feedback-erzeugung verzichten zu müssen.

Nachdem sich die Kernfunktionalität des Editors und der Feedback-erzeugung als vollständig herausgestellt hat, können in Zukunft nun auch die als weniger dringend eingestuften Anforderungen angegangen werden. Dies betrifft insbesondere die Usability, bei der die Auswahl von Orbitalen oder das Setzen von Elektronen per Mausklick in der Diagrammfläche ermöglicht werden könnte. Ferner könnten die beiden bisherigen Eingabefelder durch sogenannte „Spinner“ ersetzt werden, wodurch die Notwendigkeit einer Tastatureingabe gänzlich entfallen würde. Ferner kann die Eingabe im Editor durch eine Ausgabe in Form einer 3D-Simulation ergänzt werden, die eine räumliche Visualisierung der eingetragenen Orbitale anbietet und dadurch eine optische Verifikation der Lösung ermöglicht. Vergleichbare Darstellungen sind den Studierenden aus anderen Aufgabentypen bereits bekannt. Ähnlich wie bei der Zuweisung eines

plausiblen Energieniveaus zu fälschlicherweise eingetragenen Orbitalen muss dabei auch für die 3D-Darstellung eine optische Lösung für diese Orbitale gefunden werden, sofern die Darstellung nicht nur für korrekte Lösungen erzeugt werden soll.

Aus didaktischer Sicht ist eine Erweiterung des Editors zur Abdeckung weiterer Aufgabentypen möglich. Insbesondere können Orbitalschemata nicht nur für einzelne Atome, sondern auch für Moleküle angegeben werden. Der Editor und die dahinterstehende Datenstruktur müssen dazu so erweitert werden, dass mehrere Atome gleichzeitig abgebildet und bearbeitet werden können, was weitere Anforderungen hinsichtlich der Usability, des automatischen Layouts und der automatischen Feedbackerzeugung impliziert.

**Danksagung:** Die Autoren danken Cedric Krause für die Entwicklung eines ersten Prototyps für den Editor im Rahmen seiner Bachelorarbeit.

## Literaturverzeichnis

- [As05] Ashton, H. S. et al.: Investigating the medium effect in computer-aided assessment of school Chemistry and college Computing national examinations. *British Journal of Educational Technology*, 36: S. 771-787, 2005. doi:10.1111/j.1467-8535.2005.00501.x
- [Jo15] Jobst, C.: Potenziale neuer Fragetypen für die Naturwissenschaften. In: *Grundfragen Multimedialen Lehrens und Lernens (GML<sup>2</sup> 2015)*. S. 145–152, 2015.
- [MM14] Mortimer, C. E.; Müller, U.: *Chemie: Das Basiswissen der Chemie*. 11., vollst. überarb. Aufl., Thieme, Stuttgart, 2014.
- [PBR07] Perry, S.; Bulatov, I.; Roberts, E.: The Use of E-assessment in Chemical Engineering Education. *Chemical Engineering Transactions*, 12, S. 555-560, 2007.
- [PS19] Pobel, S.; Striewe, M.: Domain-Specific Extensions for an E-Assessment System. In (Herzog, Michael A.; Kubincová, Zuzana; Han, Peng; Temperini, Marco, Hrsg.): *Advances in Web-Based Learning – ICWL 2019*. Springer International Publishing, Cham, S. 327–331, 2019.



## Usability design and evaluation for a formative assessment feedback

Florian Horn<sup>1</sup>, Daniel Schiffner<sup>2</sup>, Thorsten Gättinger<sup>3</sup>, Patrick Sacher<sup>4</sup>

**Abstract:** In a current research project we implemented an approach for delivering computer assisted adaptive testing as an additional form of formative assessment for a lecture. We primarily used these tests as a formative assessment during a “fundamentals of computer programming” course. The tests also included an individual feedback to further guide the students. To evaluate and improve the formative assessments, we performed a usability study, which was focused on the user satisfaction while taking a test and reading the feedback. The study used the user experience questionnaire. The results indicate that an adaptive assessment can provide more support, but also shows shortcomings of the current implementation.

**Keywords:** formative assessment, adaptive testing, UX, UEQ

### 1 Introduction

The rising numbers of students at German universities makes a strongly digitalised lecture style and a focus on self-guided learning a necessity. A part of these requirements can be approached by implementing a computer-assisted adaptive test (CAT) as a formal assessment. A CAT is a test that adapts itself to the users’ performance, delivering harder questions to users who perform well, as well as easier questions to users who perform badly.

As research shows, an elaborate feedback is crucial for students trying to estimate their learning progress (see [WB96] and [Wa08]). Furthermore, it has to be ensured that the students’ user experience is not a hindrance to their learning process.

In this work we evaluate an assessment used in a beginners lecture for programming. We explain our methodology, present our results that were gathered in a comparison study.

---

<sup>1</sup> Goethe University, studiumdigitale, Robert-Mayer Str.10, 60325 Frankfurt, horn@studiumdigitale.uni-frankfurt.de

<sup>2</sup> DIPF | Leibniz Institute for Research and Information in Education, IZB, Rostocker Straße 6, 60323 Frankfurt, schiffner@dipf.de

<sup>3</sup> Goethe University, studiumdigitale, Robert-Mayer Str.10, 60325 Frankfurt, gattinger@studiumdigitale.uni-frankfurt.de

<sup>4</sup> Goethe University, studiumdigitale, Robert-Mayer Str.10, 60325 Frankfurt, sacher@studiumdigitale.uni-frankfurt.de

## 2 Related Work

There are different kinds of formative assessments. The survey [MY17] illustrates several, typically used types: Simple multiple choice quizzes, one-minute papers, ePortfolios, student response systems and many other web 2.0 tools.

Furthermore, the survey mentions several implemented formative assessment systems and their usage. Klecker used the learning management system (LMS) Blackboard in 2003 to deliver online multiple choice quizzes to students. These quizzes were instantly checked for correctness and feedback was given immediately [KI07]. A different approach is taken by CompAssess. CompAssess is a tool for customized assessments with Microsoft Office programs and was evaluated by Brink and Lautenbach [BL11]. It was positively reviewed but had some technical flaws. With PsyCAL (Psychology Computer-Assisted Learning) Buchanan developed a system for multiple choice questions with an instant feedback for wrong choices [Bu00]. The feedback contains excerpts from textbooks to help students understand their wrong answers. For post-graduate students in engineering Burrow et al reviewed three different systems for online formative assessments. Of these, TRIADS (Tripartite Interactive Assessment Delivery Tool) was the best-evaluated system [Bu05]. It may contain up to 40 different interactive questions and provides the correct solutions, as well as the required information, as feedback. The systems evaluation determined it to not be user friendly, but superior to several of its competitors. We used a similar approach of providing students with literature recommendations to revise after the assessment. Rather than providing text excerpts we provided citations.

Other approaches to formative assessments include curriculum-based measurements like the platform quop [Sc19a]. The target audience are pupils from first to sixth grade and the assessed skills are reading, English language and mathematics, starting from fifth grade. After several successful pilot projects this platform was used by over 2000 teachers and 30000 pupils in Hesse [Sc19a]. Lai et al did a user experience study of a web-based formative assessment system for English proficiency [LCC17]. The participants were 28 college students. Their work showed that, given several feedback options in a learning environment (immediate feedback for single questions, delayed feedback for a group questions or immediate feedback for several questions), the option with the highest usability rating was to give a collected feedback for a set of questions. We also used this approach in our formative assessment. The User Experience Questionnaire [LHS08], a questionnaire consisting of 26 Likert scale questions designed to measure the subjective user experience of a software product. The questions consist of opposites and users are tasked with deciding which is more fitting for the software, e.g.: whether the software is more “predictable” or “unpredictable”. These 26 questions are then grouped into 6 categories, each summarising a key aspect of the user experience: attractiveness, perspicuity, efficiency, dependability, stimulation and novelty. For each of these categories the means as well as the standard deviations were calculated and the results compared to the UEQ Benchmark data [SHT17], which consists of 401 studies from a broad range of software products. We performed our usability study using the UEQ.

### 3 Implementation

Our formative assessment is implemented as a web based training and distributed to students as a SCORM module [SC19b]. Each assessment consists of 15 questions selected from a question pool. The questions were calibrated using the item response theory [Lo80]. The adaptive algorithm targets for a difficulty that allows a student to answer half of the questions correctly.

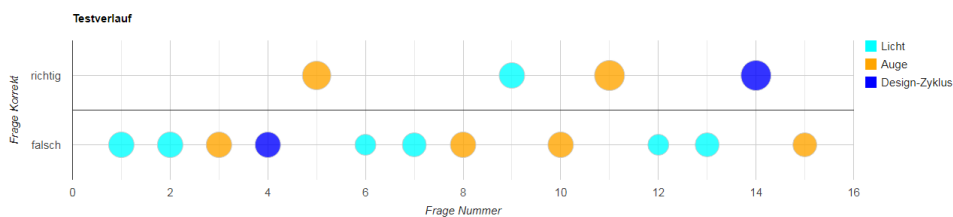


Fig. 1: The visualisation for used in feedback type B. It encodes which questions the student answered correctly (top or bottom), which subject it belongs to (colour) and the difficulty (size, the larger the easier)

Each assessment closes with a feedback presenting an overview of the performance. The basic version (type A) contains only textual feedback, including several performance indicators, a likelihood to solve other questions and literature recommendations. These recommendations varied with the students performance. A struggling student was given basic literature for the lecture, while a competent student received literature for further reading. The feedback does not include results of individual questions. We restrained from giving feedback for individual questions due to a conflict of interest in our design. Originally, we planned to use the assessment both formatively, as well as summatively. Thus, we decided not to leak solutions to questions directly.

We further enhanced the feedback using an additional visualisation (type B), which displays the test run. Correct and incorrect questions are displayed separately, the subject of the question is displayed and the relative difficulty of the question is encoded as well, as explained in figure 1.

We choose to evaluate the effect of including a visualisation in the feedback, since our assessment tool has a basic design and the first informal feedback of students and colleagues indicated that the visual stimulation was low. In addition we were asked for more granular feedback about single questions.

## 4 Evaluation

We conducted a two-fold usability study of our assessment. First, we questioned our students using short questionnaires and open feedback. Secondly, we performed a comparison study of the feedback types with different students. The latter study utilised the user experience questionnaire (UEQ) to get quantitative data. We split our assessment and its feedback to ensure that the evaluation of our feedback prototypes is independent.

Each session was planned for 15 minutes. Students were presented with a short version of a formal assessment, consisting only of 4 base questions types. Afterwards, the first UEQ had to be filled. The objective of this UEQ was to assess the general performance of the assessment tool, without the feedback, and the results are given in 3. Then they are randomly sent to one of feedback versions. The feedback is shown based on a simulated test. Under the premise that it was their feedback, the second UEQ has been filled.

## 5 Results

The result of the open evaluation (N=569) was largely considered a net positive. Students enjoyed being kept in flow, making the quizzes equally challenging while still improving their knowledge.

There were two points of critique. The major one was the feedback, especially the inability to tell the correctness of the answers. Students consider it a hindrance when using the formative assessment to study. The literature recommendations were criticised as well, as they were to unspecific.

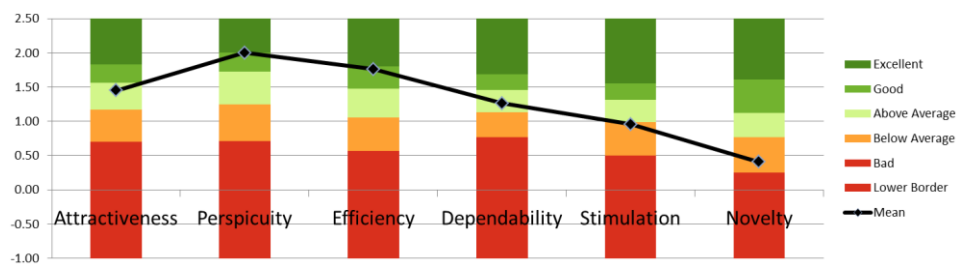


Fig. 2: UEQ Benchmark Results. The mean line illustrates how our software, i.e. the formal assessment tool, performed compared to the 401 existing benchmarks in the UEQ Benchmark, illustrated via the bar colours [SHT17]. N=85

The usability study (N=85) was evaluated using tools supplied with the UEQ [Us20]. As seen in figure 2, we performed excellently in perspicuity, good in efficiency, above average in attractiveness and dependability and below average in stimulation and novelty.



Our main goal was to design a simple and easy to use assessment. This is reflected by our performance in perspicuity, efficiency, attractiveness and dependability. As our design is very basic, our performance in stimulation and novelty is low.

A comparison of the feedback types is illustrated in figure 3 using the UEQs major scales.

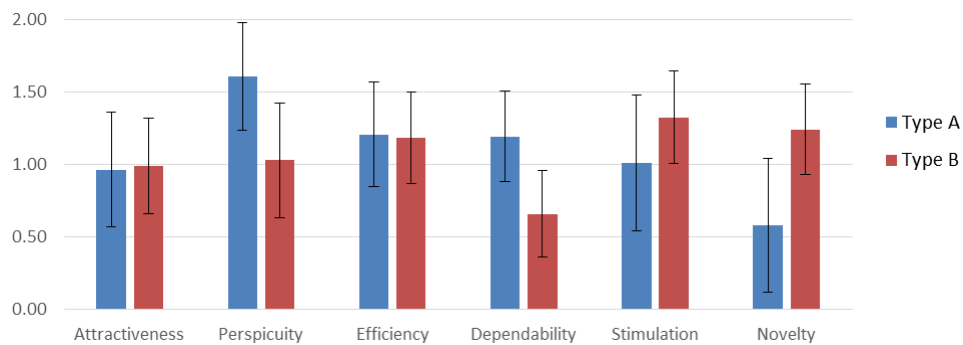


Fig. 3: Results of the A-B comparison study, Type A is the feedback iteration that has been shipped to programming students. Type B is the feedback containing information about single correct answers and the visualisation. N=85, 43 participants received Type A, 42 received Type B.

A Student's T-Test, for  $\alpha = 0.05$ , showed that the changes in perspicuity, dependability and novelty were significant. The changes in perspicuity and dependability were unexpected and contrary to the design goal we had in mind. Also, students are rarely confronted with bubblecharts, which improves the novelty but also decreases the readability of the visualisation. Similar results can be found in [BBG18]. We therefore plan to simplify the visualization.

## 6 Conclusion and Future Work

Our results indicate that the user experience of a formal assessment and its feedback are equally important and should be evaluated separately.

While being considered helpful, we also discovered several challenges when creating feedback. Our visualisation decreased the overall user satisfaction but increased the novelty.



We plan to enhance the feedback with a layered visualisation to allow for simpler representations while still providing more insights. In addition we plan to only use our assessment as a formative assessment in the future and thus will be able to give further guidance, concerning questions and answers, to students.

## Bibliography

- [BBG18] Bull, S.; Brusilovsky, P.; Guerra, J.: Which Learning Visualisations to Offer Students?: 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Leeds, UK, September 3-5, 2018, Proceedings, pp. 524–530, 01 2018.
- [BL11] Brink, R.; Lautenbach, G.: Electronic assessment in higher education. *Educational Studies*, 37(5):503–512, 2011.
- [Bu00] Buchanan, T.: The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning*, 16(3):193–200, 2000.
- [Bu05] Burrow, M.; Evdorides, H.; Hallam, B.; Freer-Hewish, R.: Developing formative assessments for postgraduate students in engineering. *European Journal of Engineering Education*, 30(2):255–263, 2005.
- [Kl07] Klecker, B. M.: The impact of formative feedback on student learning in an online classroom. *Journal of Instructional Psychology*, 34(3):3, 2007.
- [LCC17] Lai, T.; Chen, P.; Chou, C.: A user experience study of a webbased formative assessment system. In: 2017 International Conference on Applied System Innovation (ICASI). IEEE, pp. 899–902, 2017.
- [LHS08] Laugwitz, B.; Held, T.; Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Symposium of the Austrian HCI and Usability Engineering Group. Springer, pp. 63–76, 2008.
- [Lo80] Lord, F. M.: Applications of item response theory to practical testing problems. Routledge, 1980.
- [MY17] McLaughlin, T.; Yan, Z.: Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer Assisted Learning*, 33(6):562–574, 2017.
- [Sc19a] Schulen in Hessen : quop. <https://www.quop.de/de/quop-in-der-praxis/schulen-in-hessen>, 2019. Accessed: 15/03/2019.
- [SC19b] SCORM. <https://scorm.com>, 2019. Accessed: 15/03/2019.
- [SHT17] Schrepp, M.; Hinderks, A.; Thomaschewski, J.: Construction of a Benchmark for the User Experience Questionnaire (UEQ). *IJIMAI*, 4(4):40–44, 2017.
- [Us20] Usability Experience Questionnaire. <https://www.ueq-online.org/>, 2020. Accessed: 18/03/2020.
- [Wa08] Wang, T.: Web-based quiz-game-like formative assessment: Development and evaluation. *Computers & Education*, 51(3):1247–1263, 2008.
- [WB96] Wiliam, D.; Black, P.: Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British educational research journal*, 22(5):537–548, 1996.

## Biblisches Hebräisch digital lehren und lernen

### Praxiserfahrungen beim Aufbau eines Blended-Learning-Formats mit Moodle

Orlando Brix<sup>1</sup>, Susann Schober<sup>1</sup>, Leena Steinke <sup>2</sup> und Sven Strickroth <sup>3</sup>

**Abstract:** In dieser Arbeit wird vorgestellt, wie ein Sprachkurs für Biblisches Hebräisch mit E-Learning-Angeboten kombiniert werden kann. Zentrale Rolle spielen dabei interaktive und formative Assessments für Übungen und Tests zum Vokabel- und Grammatikwissen, die im veranstaltungsbegleitenden Kurs auf dem LMS Moodle angeboten werden. Ziel ist es, dass die Lernenden mit sehr heterogenem Vorwissen so selbstbestimmt, ortsunabhängig und zielgerichtet ihre Qualifikationen im Biblischen Hebräisch erwerben bzw. weiterentwickeln können. Zudem können so von der Lehrperson Problemfelder identifiziert werden, um diese in der Lehre aufzugreifen. In dieser Arbeit wird erläutert, welche Fragetypen von Moodle zur Übung von welchen Lerninhalten geeignet sind sowie welche sprachspezifischen Besonderheiten Anforderungen an die genutzten Formate, die Umsetzung und die verwendeten Tools stellen.

**Keywords:** Sprachkurs, Biblisches Hebräisch, Blended Learning, E-Assessment, Moodle

## 1 Einleitung

An der Universität Potsdam wird die Altsprache Biblisches Hebräisch aufbauend auf den Kenntnissen des Modernen Hebräisch gelehrt. Im Gegensatz zum Sprachkurs Modernes Hebräisch, in dem Lese- und Hörverständnis sowie Schreiben und Sprechen grundlegend vom Lernenden entwickelt werden, steht beim Biblischen Hebräisch das Leseverständnis im Mittelpunkt: Lernende müssen mittelschwere erzählende Texte der Hebräischen Bibel (Torah) flüssig vorlesen, die komplexen syntaktischen Strukturen verstehen und diese mithilfe eines wissenschaftlichen Wörterbuchs eigenständig erarbeiten können. Daraus folgt, dass die Syntax und die Analyse von Wortformen im Vordergrund des Sprachkurses stehen. Hervorzuheben ist, dass Hebräisch von rechts nach links geschrieben wird und das Alphabet nur aus Konsonanten besteht. Zusätzlich gibt es ein Vokalzeichensystem (Niqqud), das zum Lesen des Bibeltexts benötigt wird. Der Sprachkurs findet über ein Semester statt und hat einen Umfang von 8 SWS. Die Lehre und das Lernen sind durch Präsenzunterricht und individuelle Lernphasen gekennzeichnet. Der erfolgreiche Abschluss des Moduls und damit der beiden Kurse Modernes und Bibli-

---

<sup>1</sup> Universität Potsdam, School of Jewish Theology, Am Neuen Palais 10, 14469 Potsdam, orlando.brix@uni-potsdam.de; susann.schober@uni-potsdam.de

<sup>2</sup> Universität Potsdam, Institut für Informatik, August-Bebel-Straße 89, 14482 Potsdam, leena.steinke@uni-potsdam.de, <https://orcid.org/0000-0002-4693-9414>

<sup>3</sup> Universität Potsdam, Institut für Informatik, August-Bebel-Straße 89, 14482 Potsdam, sven.strickroth@uni-potsdam.de, <https://orcid.org/0000-0002-9647-300X>

sches Hebräisch, resultiert im Hebraicum. Lernende des Kurses Biblisches Hebräisch an der Universität Potsdam sind Studierende der Jüdischen Theologie, der Jüdischen Studien und der Religionswissenschaft. Daher gibt es Lernende im Sprachkurs, die zuvor nie mit Biblischem Hebräisch in Kontakt gekommen sind und Lernende, die bereits hebräische Gebete kennen und vertraut sind mit Texten aus der Torah. Daraus folgen sehr heterogene Bedürfnisse an die Inhalte des Kurses. Eine Über- oder Unterforderung der Lernenden kann sich negativ auf die Motivation und Leistungsbereitschaft der Lernenden auswirken. Dem Ansatz des Blended Learning folgend wird der bestehende, bisher nur zum Bereitstellen von Materialien genutzte, Moodle-Kurs um interaktive und formative Assessments erweitert. Diese Erweiterungen erfolgen im engen Austausch und auf Basis der methodologischen und didaktischen Erfahrungen der Lektorin des Sprachkurses. Die Verwirklichung des Vorhabens erfolgt im Rahmen des BMBF-geförderten Projekts „Forschen | Lernen – Digital“ (FoLD).

Eine besondere Herausforderung, digital unterstützte Lehre in der traditionsreichen Jüdischen Lernkultur am Institut der Jüdischen Theologie zu etablieren, besteht nicht nur darin, geeignete digitale Lehr- und Lernmittel zu entwickeln und umzusetzen, sondern vor allem grundsätzlich den Einsatz in der Lehre zu rechtfertigen. Denn bei manchen Lehrenden der Jüdischen Theologie ist eine grundsätzliche Ablehnung gegenüber der Digitalisierung zu beobachten; sie bestehen auf den traditionellen Methoden des Mündlichen und Schriftlichen Lehrens und Lernens. Digitale Tools scheinen da fremd. Zudem ist der sinnvolle Einsatz digitaler Tools in der Lehre für Lehrende ohne viel technisches Vorwissen mit hohem zusätzlichem Zeit- und Arbeitsaufwand verbunden. Es ist daher wichtig, aufzuzeigen, dass digitale Methoden und Werkzeuge sinnvoll und nachhaltig in der Lehre eingesetzt werden können und Lehrende und Lernende dauerhaft davon profitieren.

Im Folgenden werden ausgehend vom Stand der Forschung der entwickelte Blended-Learning-Ansatz sowie Erfahrungen aus der Implementierung im LMS Moodle dargestellt. Abschließend folgen eine kurze Zusammenfassung und ein Ausblick.

## 2 Stand der Forschung

Seltener unterrichtete Sprachen, wie das Biblische Hebräisch, profitieren besonders von Computer-Assisted-Language-Learning (CALL) [BI08]. Für Biblisches Hebräisch gibt es im deutschsprachigen Raum bereits Ansätze, das Lernangebot durch digitale Lernmittel zu erweitern. In 2006 wurde im Seminar für Judaistik an der Universität Frankfurt ein begleitender Online-Kurs zu den Sprachkursen eingeführt. Das Material dieses Kurses ist allerdings nicht öffentlich zugänglich und in das LMS OLAT eingebettet.<sup>4</sup> An der Universität Greifswald wurden E-Learning-Angebote im LMS Moodle für den Sprachkurs entwickelt, den Studierende der evangelischen Theologie besuchen [De18]. Dort wurde ein dreiteiliges Szenario, bestehend aus dem Präsenzunterricht, dem E-Learning

---

<sup>4</sup> vgl. [https://www.uni-frankfurt.de/43117276/Online\\_Begleitung\\_Sprachkurse](https://www.uni-frankfurt.de/43117276/Online_Begleitung_Sprachkurse) (letzter Abruf: 2020-06-20)

mit Moodle und einem TutorInnenprogramm entwickelt. Zudem sahen sich die Lehrenden mit einem Mangel an grammatischem Grundwissen in der deutschen Sprache konfrontiert, der die Vermittlung der Grammatik der hebräischen Sprache erschwert. Daraufhin haben sie ein Wiki für den Moodle-Kurs entwickelt, das neben syntaktischen und grammatischen Besonderheiten des Hebräischen auch elementare grammatische Begriffe des Deutschen erklärt. Weiterhin stehen ein Vokabeltrainer mit Bildschirmtastatur sowie ein Memory zum Lernen zur Verfügung [De18]. Leider war ein Wissensaustausch mit oben genannten Lehrenden nicht erfolgreich.

### 3 Ansatz für einen Blended-Learning Kurs

Für den sinnvollen Einsatz digitaler Lehr- und Lernmittel wurde sich an Robert J. Blake [B108] orientiert. Die digitalen Lehr- und Lernmittel ersetzen weder Lehrende noch greifen sie in die Lehre ein, sondern unterstützen die Lehrenden und Lernenden nachhaltig. Im Sinne des Blended-Learnings nach [Sh17] wird der Sprachkurs um interaktive Übungen im Moodle-Kurs erweitert. Die bis jetzt händisch während des Präsenzunterrichts ausgefüllten Zwischentests werden ebenfalls digitalisiert und den Lernenden nach bestimmten Lehrinhalten im Moodle-Kurs zur Verfügung gestellt. Dadurch hat die Lehrende Einblick in den Lernstand der Lernenden und kann daraufhin das Tempo der Lehre anpassen und ggf. auf Schwierigkeiten bei bestimmten Themen gezielt eingehen. Durch die Bereitstellung digitaler Lernmittel haben die Lernenden nicht nur die Möglichkeit, zeit- und ortsunabhängig zu lernen, sondern bekommen auch unmittelbares Feedback über ihren Lernstand und können zielgerichtet lernen. Die Arbeit am Bibeltext findet weiterhin im Unterricht statt, während die Lernangebote im Moodle-Kurs vor allem für das Festigen der Grammatik und des Wortschatzes genutzt wird (siehe Tab. 1). Eine unabdingbare Fähigkeit zum digitalen Lernen des Hebräischen ist der Umgang mit der Hebräischen Tastatur. Daher ist ein weiteres Ziel des Sprachkurses, dass die Lernenden sicher Hebräisch mit und ohne Niqqud an ihrem Endgerät schreiben können. Dadurch können sie die gesamte digital zugängliche Primär- und Sekundärliteratur des Forschungs- und Nachschlagefeldes des Biblischen Hebräischs nutzen. Diese Fähigkeit wird für die weiterführenden Kurse im Studiengang der Jüdischen Theologie benötigt. D. h. neben den Lernkompetenzen werden durch die Erweiterung des Sprachkurses um interaktive Lehr- und Lernmittel auch Digitalkompetenzen nach [Eh20] gefördert.

Für die Bereitstellung interaktiver Lehr- und Lernmittel wird das LMS Moodle genutzt. Dabei stehen eine Reihe unterschiedlicher Fragetypen zur Verfügung, von denen sich einige als besser geeignet für Übungen und Tests der Vokabel- und Grammatikkenntnisse im Sprachkurs Biblisches Hebräisch gezeigt haben (vgl. Tabelle 1). Maßgebliche Kriterien waren die Lesbarkeit des Hebräischen, die Realisierbarkeit von Aufgaben mit Niqqud und ein angemessenes Arbeitsaufwand-Nutzen-Verhältnis bei der Erstellung in Moodle. Wünschenswert wären spezielle Fragetypen in Tabellenform für Konjugationen von Verben.

Typ	Analoge Aufgabe	Digitales Format	Übungsinhalt
V	Vokabelübung/-test	Kurzantwort	Hebr. Wort gegeben, Abfrage möglicher Übersetzungen (Feedback: alle Übersetzungen)
VG	Verbalform bestimmen (+Übersetzung)	Lückentext (Cloze)	Konjugationstabelle mit Lücken
G	Verbalform bilden	Kurzantwort	Wortstamm (Shoresh), Stammvokal & Übersetzung vorgegeben; Eingabe: Hebr. Form
G	Konsonantenverdoppler (Dagesch forte und lene)	Zuordnungsfrage	Drop-Down-Menü zur Auswahl des richtigen Dagesch
G	Pronominales Suffix von Verbal-/Nominalformen bestimmen	Drag'n'Drop (Text)	Dt. Übersetzung in einer Spalte, Zuordnung passender hebr. Form (vgl. Abb. 1)
G	- " -	Lückentext (Cloze)	Analyse hebr. Form in Tabelle (vgl. Verbalform bestimmen)
G	Vokalzeichen (Qamatz qatan)	Zuordnungsfrage	Hebr. Worte mit Drop-Down-Menü, ob Qamatz als „a“ oder „o“ gelesen wird
G	Paare aus ähnlichen Verbalformen finden	Drag'n'Drop (Text)	Spalte mit Verbalformen, passende hebr. Form ist gesucht

Tab. 1: Umsetzungsmöglichkeiten für bestehende Aufgabenformen; V=Vokabeln, G=Grammatik

Als eine besondere Herausforderung hat sich die Eingabe, Darstellung und Antwortüberprüfung in Moodle herausgestellt. Für die Erstellung von Übungen, speziell bei der Erstellung von Lückentexten, muss zwischen den Tastaturbelegungen Deutsch und Hebräisch mit Vokalzeichen gewechselt werden. Während eine Mischung aus hebräischer und deutscher Texteingabe im Fragetext zufriedenstellend wiedergegeben wird, ergeben sich in den Antwortfeldern durch die unterschiedlichen Schreibrichtungen (innerhalb einer Zeile) bei der Eingabe Fehlerquellen. Bei der Erstellung der Tests muss also darauf geachtet werden, dass die Antworten in je einer einzigen Schrift/Sprache einzugeben sind. Ist weiterhin eine einheitliche Schriftgröße für lateinische oder altgriechische Buchstaben völlig ausreichend, sind die Vokalzeichen nur sehr schwer zu erkennen (vgl. Abb. 1). Bei Aufgabentypen mit WYSIWYG-Editor lässt sich eine größere Schrift wählen, jedoch sind allgemein in Antwortfeldern die Schriftgrößen durch das Moodle-Template vorgegeben und lassen sich nur im Browser durch Erhöhung des Zoom-Levels verändern. Auch die Eingabe des Niqud über diakritische Zeichen stellt eine Herausforderung dar, da mehrere Zeichen (speziell in Moodle) ähnlich aussehen, jedoch eine andere Bedeutung bzw. Unicode-Repräsentation haben: z. B. falsch מלך bzw. מלך (Sin

Punkt; U+05C2) vs. richtig  $\text{חָלָם}$  (Cholam; U+05B9) bzw. Verwendung von  $\text{ָ}$  (Qamatz qatan; U+05C7) anstatt  $\text{ַ}$  (Qamatz gadol; U+05B8). Diese Probleme treten insbesondere bei mobilen Geräten auf, bei denen externe Apps für die Eingabe hebräischer Zeichen notwendig sind. Die Apps bieten, anders als Windows und MacOS über die Tastatur, eine größere Auswahl bereits komponierter Glyphen an, was Studierende zu überfordern scheint. Verschiedene Eingabereihenfolgen der diakritischen Zeichen bzw. äquivalente Glyphen werden von Moodle zuverlässig erkannt (z. B.  $\text{חָ}$  (U+FB2D) =  $\text{ח}$  (U+05E9) +  $\text{ָ}$  (U+05BC) +  $\text{ָ}$  (U+05C2) =  $\text{ח}$  +  $\text{ָ}$  +  $\text{ָ}$ ; auch wenn der Unicode-Standard die Reihenfolge von innen vom Basiszeichen nach außen vorsieht<sup>5</sup>). Bei der automatischen Feedbackgabe müssen daher konkrete Hinweise zur richtigen Verwendung der entsprechenden Niquid gegeben, evtl. mehrere Varianten als Lösung hinterlegt und Tutorials für die richtige Eingabe bereitgestellt oder ggf. auch andere Aufgabentypen genutzt werden.

**Frage 1**  
Bisher nicht beantwortet  
Erreichbare Punkte: 1,00  
Frage markieren  
Frage bearbeiten

**Füge die passende Form hinzu!**

dein (m) Wort	<input type="text"/>				
ihr (f.sg.) Wort	<input type="text"/>				
mein Wort	<input type="text"/>	דברך	דברכם	דברי	דברה
euer (m) Wort	<input type="text"/>	דברכן	דברו	דברנו	דברו
unser Wort	<input type="text"/>	דברך	דברם		

Abb. 1: Beispiel der Umsetzung einer Übungsaufgabe mittels Drag'n'Drop in Moodle

## 4 Zusammenfassung und Ausblick

In diesem Beitrag wird beschrieben, wie der Sprachkurs Biblisches Hebräisch an der Universität Potsdam um interaktive Übungen und Tests über das LMS Moodle erweitert wird. Ziel ist es, zusätzliche Lernangebote anzubieten und den Lehrenden Einblick in den Lernstand zu ermöglichen. Darüber hinaus werden dadurch Digitalkompetenzen bei den Beteiligten aufgebaut und eine langfristige Nutzung der Materialien angestrebt. Es hat sich gezeigt, dass die Schreibrichtung und Niquid besondere Anforderungen an die Darstellung in Moodle als auch an das Design der Aufgaben stellen. Spezielle Unterstützung für das Erkennen ähnlicher diakritischer Zeichen oder ein Aufgabentyp für z. B. Konjugationstabellen wären wünschenswert. Spannend ist vor allem auch die Einführung spielerischerer Übungsformate, wie z. B. das Memory in [De18]. Die Ergebnisse sind auf Sprachkurse übertragbar, die eine Kombination diakritischer Zeichen benötigen bzw. eine Schreibrichtung von rechts nach links haben (z. B. Arabisch).

Ein nächster Schritt besteht in der Evaluation des entwickelten Ansatzes. Dazu werden die Teilnehmenden am Ende des Semesters befragt und die Anzahl der Teilnehmenden sowie Prüfungsergebnisse des aktuellen Semesters mit vorherigen Durchläufen des Kur-

<sup>5</sup> <https://www.unicode.org/versions/Unicode12.0.0/ch02.pdf>, Kapitel 2.11 (letzter Abruf: 2020-06-20)

ses verglichen. So sollen Erkenntnisse über den Nutzen der E-Learning-Angebote gewonnen werden. Zudem wird zur Untersuchung von Kompetenzzuwächsen ein Test zur Kompetenzmessung am Anfang und Ende des Semesters durchgeführt. Erste Ergebnisse der Vorstudie sollen nach dem Sommersemester 2020 vorliegen. Des Weiteren bleibt festzustellen, ob sich die Präsenz der Studierenden in den Hebräisch-Kursen durch die Einführung digitaler Lernhilfen verändert. Die Präsenz in den Hebräisch-Kursen war in den vergangenen Semestern sehr hoch, obwohl es keine Anwesenheitspflicht gibt.

Auf Grund der aktuellen Maßnahmen zur Pandemie-Eingrenzung wird der Kurs noch weiter zu einem reinen online Kurs ausgebaut. Synchron Meetings bleiben für den Sprachkurs unerlässlich. Das erweiterte Konzept sieht vor, die synchronen Unterrichts-Meetings, in mehreren verkürzten Blöcken über eine Videokonferenzplattform zu halten und diese durch Übungsblöcke zu ergänzen. In den Übungsblöcken erarbeiten sich die Lernenden Inhalte und stellen ihre Ergebnisse dann den anderen in synchronen Meetings vor. Bei technischen Problemen oder Fragestellungen steht den Studierenden ein Moodle-Forum zur Verfügung, das von wiss. Hilfskräften betreut wird. Zudem werden die Angebote im Moodle-Kurs weiter ausgebaut: Es gibt extra Übungen, um das Schreiben mit der hebräischen Tastatur mit und ohne Niqqud zu lernen, weitere Übungen zum Vokabeltraining und Audio-Dateien zur Unterstützung des Lese-Lernens. Nach der Hälfte des Semesters zeigt sich, dass das Hilfe-Forum von den Lernenden positiv angenommen und genutzt wird. Bisher scheint es, als ob die E-Assessments, die das Schreiben mit der hebräischen Tastatur benötigen, weniger genutzt werden – die Übungen zum Vokabelwissen mit Drag'n'Drop dagegen deutlich häufiger. Am Ende des Semesters muss evaluiert werden, welche Übungsformate von den Lernenden als sinnvolle Unterstützung in ihrem Lernprozess wahrgenommen werden. Die Lektorin wünscht sich weitere E-Assessments. Abschließend ist festzustellen, dass im digitalen Semester die Akzeptanz gegenüber digitaler Tools in der Lehre angestiegen ist.

**Danksagung:** Das Projekt FoLD wird vom BMBF unter dem Kennzeichen 16DHB3018 gefördert. Dank gilt auch den wiss. Hilfskräften des FoLD-Projekts, durch deren technische Unterstützung die Weiterführung der Lehre im digitalen Semester erst möglich ist.

## Literaturverzeichnis

- [Bl08] Blake, R. J.: *Brave New Digital Classroom. Technology and Foreign Language Learning*, Georgetown University Press, Washington, D.C., 2008.
- [De18] Degwitz, F.: *Reform des Spracherwerbs Hebräisch, Griechisch, Latein*. In: *Greifswalder Beiträge zur Hochschullehre – Erleichterung der Studieneingangsphase*. Ausgabe 9, Oktober 2018.
- [Eh20] Ehlers, U. D.: *Future Skills: Lernen der Zukunft – Hochschule der Zukunft*. Springer-Verlag, 2020.
- [Sh17] Sharma, P.: *Blended learning design and practice*. In: *Digital Language Learning and Teaching*. Routledge, S. 167–178, 2017.



## Automatisierte Generierung von Automaten und automatenbasierten Aufgaben

Sven Judel<sup>1</sup>, Timo Bergerbusch<sup>2</sup>, Ulrik Schroeder<sup>1</sup>

**Abstract:** Dieser Beitrag stellt das Tool *Automata Task Random Generator (AuTaRG)* vor, das für die Bachelorvorlesung *Formale Systeme, Automaten, Prozesse* der RWTH Aachen automatenbasierte Aufgaben generiert. Es wird präsentiert, welche automatenbasierte Aufgabentypen behandelt werden, was das jeweilige Lernziel des Aufgabentyps ist, wie Eigenschaften und Zahlenwerte, die ein Automat besitzen muss um für die einzelnen Aufgabentypen geeignet zu sein, ermittelt wurden und wie solche Automaten generiert werden können. In einer Evaluation wurde erwiesen, dass durch die Nutzung des Tools Zeit bei der Generierung der Aufgaben und der Überführung in LaTeX-Code gespart wird und eine, im Vergleich zur händischen Erstellung, größere Aufgabenvariation und –menge generiert werden kann. Dies führt dazu, dass Assistenten weniger Aufwand im Übungsbetrieb betreiben müssen und Studierende mehr Aufgaben zum Üben erhalten.

**Keywords:** Automatisierte Generierung, Automaten, Automatentheorie

### 1 Einleitung

Die händische Erstellung von Übungsaufgaben und ihrer Lösungen ist zeitintensiv und birgt die Gefahr von Fehlern. Daher werden oft alte Aufgaben übernommen oder angepasst. Die automatisierte Generierung von Aufgaben und deren Lösungen soll die Erstellung beschleunigen, auf Korrektheit prüfen und eine größere Aufgabenvariation und –menge ermöglichen. [Ku20] Für die Umsetzung einer automatisierten Aufgabengenerierung wurden automatenbasierte Aufgaben der Vorlesung *Formale Systeme, Automaten, Prozesse (FoSAP)* an der RWTH Aachen betrachtet. Diese Vorlesung ist ein Pflichtmodul des Informatik-Bachelorstudiums. Die Ausgabe der wöchentlichen Aufgaben erfolgt durch den Upload eines PDFs in den Moodle-Lernraum des Kurses.

Dieser Beitrag gibt eine Antwort auf die Frage: *Wie können geeignete automatenbasierte Aufgaben für die FoSAP Vorlesung generiert werden?* Zur Beantwortung dieser Forschungsfrage wurden vier Teilforschungsfragen vorangestellt:

1. Welche automatenbasierten Aufgabentypen gibt es?
2. Wann ist ein Automat für einen Aufgabentypen verwendbar?

---

<sup>1</sup> RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstr. 55, 52074 Aachen, {judel, schroeder}@informatik.rwth-aachen.de

<sup>2</sup> RWTH Aachen, Ahornstr. 55, 52074 Aachen, timo.bergerbusch@rwth-aachen.de

3. Was definiert eine Aufgabe als geeignet?
4. Wie können verwendbare Automaten für einen Aufgabentypen generiert werden?

Hier wird zwischen *Aufgabentyp*, einem Konzept (z. B. „Addition“), und *Aufgabe*, einer konkreten Instanz eines Aufgabentyps (z. B. „Berechne  $2 + 2$ “) unterschieden.

Die Fragen 1 bis 3 werden in Kapitel 2 beantwortet, während Kapitel 3 das Konzept des erarbeiteten Generators als Antwort für Frage 4 präsentiert. In Kapitel 4 wird die Umsetzung des Generators als Onlinetool vorgestellt und in Kapitel 5 die Evaluation der Performance und Bedienbarkeit. Kapitel 6 fasst den Beitrag zusammen und stellt mögliche Weiterentwicklung für die Generierung von Aufgaben sowie der Nutzung dieser für automatisierte Bewertungen vor.

## 2 Aufgabenanalyse

Die Beantwortung der ersten drei Teilforschungsfragen beruht auf der Analyse der verfügbaren Übungsmaterialien. Die Aufgaben der Jahre 2015 bis 2019 wurden gesichtet und acht automatenbasierte Aufgabentypen identifiziert. Die verwendeten Automaten wurden auf ihre Eigenschaften und Zahlenwerte hin untersucht, um eine Eignungsklassifikation von Aufgaben für Aufgabentypen zu erstellen.

### 2.1 Gruppierung der existenten Aufgaben

Tab. 1 gibt die acht identifizierten Aufgabentypen, als Antwort auf Teilforschungsfrage 1, wieder. Die Gruppierung basiert auf den automatenbasierten Konzepten der Formale Systeme, Automaten, Prozesse (FoSAP) Vorlesung. Beispielhaft werden drei dieser Aufgabentypen zusammen mit ihrem Lernziel beschrieben.

Einführung	Zustandsäquivalenz	Produktautomat	NFA-Pfade
Spracherkennung	Blockverfeinerung	Potenzmengenkonstruktion	$\epsilon$ -NFA

Tab. 1: Die acht identifizierten Aufgabentypen

Als Einstieg in das Thema *Automaten*, mit dem Ziel die Studierenden mit den Grundlagen vertraut zu machen, wird der Aufgabentyp *Einführung* ausgegeben. Zu einem gegebenen Deterministischen Endlichen Automaten (*deterministic finite automata - DFA*)  $\mathcal{A}$ , soll das äquivalente 5-Tupel aufgeschrieben werden. Anschließend soll für die Wörter einer bestimmten Länge angegeben werden, ob  $\mathcal{A}$  sie akzeptiert. Abschließend soll eine umgangssprachliche Beschreibung aller von  $\mathcal{A}$  akzeptierten Wörter angegeben werden. Ziel dieses Typs ist die Studierenden mit den Grundlagen vertraut zu machen.

In Aufgaben vom Typ *Produktautomat* soll das Produkt  $\mathcal{P}$  von zwei Automaten  $\mathcal{A}$  und  $\mathcal{B}$  konstruiert werden, sodass  $\mathcal{P}$  die Vereinigung der von  $\mathcal{A}$  und  $\mathcal{B}$  akzeptierten Wörter selbst

akzeptiert ( $\mathcal{L}_P = \mathcal{L}_A \cup \mathcal{L}_B$ ). Dadurch soll die Übertragbarkeit von Mengenoperationen über Sprachen auf Automaten verdeutlicht werden.

Mit dem Aufgabentyp *NFA-Pfade* wird das Thema Nichtdeterminismus behandelt. Zu einem gegebenen Nichtdeterministischen Endlichen Automaten (*non-deterministic finite automata - NFA*)  $\mathcal{A}$  und einem Wort  $\omega$  sollen alle möglichen Pfade für  $\omega$  in  $\mathcal{A}$ , die Menge der mit  $\omega$  in  $\mathcal{A}$  erreichbaren Zustände und ob  $\mathcal{A}$   $\omega$  akzeptiert angegeben werden. Ziel ist die Auseinandersetzung mit und das Verständnis von Nichtdeterminismus.

## 2.2 Analyse der Automateneigenschaften

Für die Klassifizierung eines Automaten hinsichtlich der Eignung für einen Aufgabentyp und deren Aufgaben, wurden die Automaten der vorhandenen Aufgaben auf charakterisierende Eigenschaften (für die Aufgabentypen) und Zahlenwerte (für die Aufgaben) hin untersucht.

Die betrachteten Eigenschaften waren *zugänglich*, *co-zugänglich*, *akzeptierend*, *vollständig* und *deterministisch*. Diese fünf Eigenschaften reichten aus, um die Automaten im gegebenen Kontext zu klassifizieren.

In Tab. 2 ist die Analyse der Eigenschaften der drei verfügbaren Automaten zum Aufgabentyp *Potenzmengenkonstruktion* gegeben. Zur besseren Übersicht wurde zusätzlich zur Deterministisch-Spalte eine Nichtdeterministisch-Spalte hinzugefügt. Die Eigenschaften der drei Automaten sind in je einer Zeile gegeben. Das Vorhandensein einer Eigenschaft wird für jede Spalte aufsummiert und der relative Wert (Prozentwert) für die Gesamtmenge der Automaten berechnet. Basierend auf diesem Prozentwert wird eine Eigenschaft als *unzulässig* (0%), *vernachlässigbar* ( $0 < x \leq 50\%$ ), *gewünscht* ( $50\% < x < 100\%$ ) oder *notwendig* (100%) klassifiziert.

	Zugänglich	Co-zugänglich	Akzeptierend	Vollständig	Deterministisch	Nicht-deterministisch
1	✓	✓	✓	✗	✗	✓
2	✗	✓	✓	✗	✗	✓
3	✗	✓	✓	✗	✗	✓
Summe	1	3	3	0	0	3
Prozentwert	33%	100%	100%	0%	0%	100%
Ergebnis	Vernachlässigbar	Notwendig	Notwendig	Unzulässig	Unzulässig	Notwendig

Tab. 2: Eigenschaften der verfügbaren Automaten zum Aufgabentyp Potenzmengenkonstruktion

Zusätzlich zu den Eigenschaften, die ein Automat hat oder nicht, wurden die Zahlenwerte *Anzahl Zustände* ( $n$ ), *Länge des Alphabets* ( $k$ ), *Übergangsdichte* und *absolute* und *relative Anzahl Finalzustände* für jeden Automaten bestimmt. Tab. 3 gibt die Ergebnisse der Zahlenwertanalyse für die drei Automaten der Aufgaben zur Potenzmengenkonstruktion

wieder. Die Werte der einzelnen Automaten sind in je einer Zeile gegeben. Zu jedem Zahlenwert wurden das Minimum, das Maximum und der Durchschnitt berechnet.

	<b>n</b>	<b>k</b>	<b>Übergangsdichte</b>	<b>Finalzustände (absolut)</b>	<b>Finalzustände (relativ)</b>
1	4	3	0,1667	1	25%
2	5	3	0,1333	3	60%
3	5	3	0,16	1	20%
Minimum	4	3	0,1333	1	20%
Durchschnitt	4,6667	3	0,1533	1,6667	35%
Maximum	5	3	0,1667	3	60%

Tab. 3: Zahlenwerte der verfügbaren Automaten zum Aufgabentyp Potenzmengenkonstruktion

Mit diesen Ergebnissen leiten sich die notwendigen Eigenschaften (die zu erfüllen sind) und den unzulässigen Eigenschaften (die nicht gegeben sein dürfen) ab, welche die Eignung eines Automaten für einen Aufgabentypen festlegen. Die initialen Eignungsherleitungen für alle acht Aufgabentypen sind die Antwort auf Teilforschungsfrage 2. Diese sollen mit dem Feedback aus der Nutzung des Tools im Lehrbetrieb kontinuierlich angepasst werden.

### 2.3 Aufgabenkriterien

Eine Aufgabe ist für einen Aufgabentyp geeignet, wenn durch die Bearbeitung das Lernziel erreicht wird. Dies stellt einige Bedingungen (*Kriterien*) an die Aufgabenstellung und Lösungen. So müssen z. B. Aufgaben vom Typ *NFA-Pfad* über einen Automaten verfügen, der für das gegebene Wort mindestens zwei verschiedene Pfade besitzt. Die Lösung einer Aufgabe zum Produktautomat soll weniger als 12 Zustände haben, um noch gut per Hand gelöst und gezeichnet werden zu können, und gleichzeitig mehr Zustände als der Automat der Aufgabenstellung. Der Rahmen von Automaten, deren Größe für den Übungsbetrieb geeignet sind, wird weiter durch die erhobenen Intervalle (Minimum bis Maximum) für jeden Zahlenwert definiert.

Für jeden Aufgabentyp gibt die Kombination der Zahlenwertintervalle und Kriterien die Definition einer geeigneten Aufgabe an und somit die Antwort auf Teilforschungsfrage 3.

## 3 Automatengenerator

Nach der Analyse der verfügbaren Aufgaben und der Beantwortung der ersten drei Fragen werden nun die Ergebnisse genutzt um einen Generator für Automaten zu konzipieren. Dabei stellen sich die Fragen nach einem geeigneten Generierungsverfahren und einer geeigneten Generierungsstrategie.

### 3.1 Generierungsverfahren

Zunächst wurde nach vorhandenen Generatoren gesucht, die zufällige Automaten generieren. Die Forderung nach der zufälligen Generierung verhindert die Notwendigkeit einen Automaten vorzugeben, was zu der oben beschriebenen Wiederverwendung oder Anpassung alter Automaten führen kann.

Die drei vielversprechendsten der gefundenen Generatoren von NFAs (ein DFA ist per Definition auch ein NFA), *Bit Streaming*, *Leslie Generator* und *FIFA Generator* wurden näher betrachtet. Beim *Bit Streaming* wird eine Bit-Sequenz der Länge  $n^2 \times k$  generiert, in welcher eine 1 als Transition zwischen zwei Zuständen interpretiert wird [Ch04]. Ein Beispiel ist in Abb. 1 gegeben.

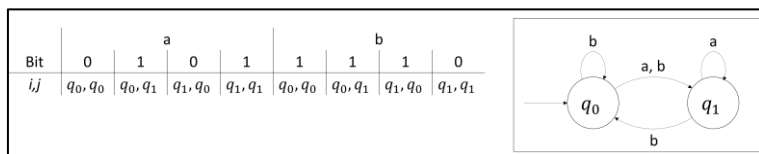


Abb. 1: Darstellung eines Automaten als Bitstream und als Transitionsdiagramm

Ein Nachteil vom Bit Streaming ist die Möglichkeit unzusammenhängende Automaten zu generieren. Dies wird vom *Leslie Generator* durch die Nutzung von Bit Streaming in Kombination mit einer zusätzlichen Bedingung und dem Parameter *Übergangsdichte* umgangen. Zunächst wird ein Automat als zufällig verbundene Struktur generiert, sodass es keine isolierten Zustände gibt, und anschließend Transitionen hinzugefügt, um die gewünschte Dichte zu erreichen. [Le95]

Die Notwendigkeit einen weiteren Parameter angeben zu müssen wirkt aber der zufälligen Generation entgegen. Dem *FIFA Generator* wiederum reichen die Parameter  $n$  und  $k$ . Ein *FIFA* (*forward injective finite automata*) ist ein NFA, der u. a. initial verbunden ist und durch einen kanonischen String repräsentiert werden kann. [Fe18]

Beginnend mit dem Startzustand wird angegeben, mit welchen Buchstaben des Alphabets auf bisher gefundene und auf neu entdeckte Zustände verwiesen wird. Die Buchstaben des Alphabets werden durch Nummern als Indizes eines Arrays aller Buchstabenkombinationen inklusive *keine Verbindung* dargestellt. Für das Alphabet  $\Sigma = \{a, b\}$  wäre dies das Array [„keine Verbindung“, „a“, „b“, „a, b“]. Abb. 2 gibt einen Automaten mit diesem Alphabet und  $n = 5$  zusammen mit seiner Darstellung durch den kanonischen String an. Der Startzustand ist A. Er besitzt keine Transition auf sich selbst, dargestellt durch die 0. Von A aus erreicht man die Zustände B und C mit den Buchstaben a (Index 1) bzw. b (Index 2). Die Suche nach Folgezuständen erfolgt durch eine Breitensuche nach der Reihenfolge der Buchstaben im Array. So werden zuerst die Folgezustände für a, dann für b und abschließend für a, b gesucht. Der Zustand B hat keine Transition zurück auf A, aber für die Buchstaben a und b (Index 3) eine Transition auf sich selbst. Er hat ebenfalls keine Transition zu C und auch keine Transitionen auf neue Zustände. Die Zustände C, D

und  $E$  werden analog notiert. Die letzte Liste gibt binär an, welcher der gefundenen Zustände ein Endzustand ist. In dem gegebenen Beispiel ist dies Zustand  $D$ , der als viertes gefunden wurde, sodass nur an der vierten Stelle des Arrays eine  $1$  steht.

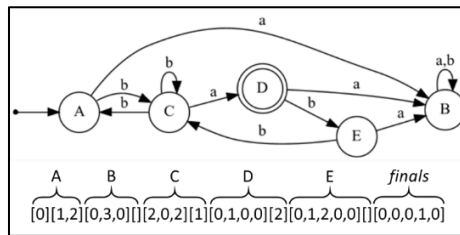


Abb. 2: Ein Automat als Transitionsdiagramm und als kanonischer String

Der FIFA Generator erstellt einen Automaten mit  $n$  Zuständen durch Breitensuche mit der zufälligen Auswahl von Buchstaben für jede Zustandskombination, sodass Verbundenheit garantiert wird. So lassen sich für beispielweise  $n=k=2$  ohne Berücksichtigung von Endzuständen 192 Automaten generieren, mit der Berücksichtigung von Endzuständen 768. Für  $n=5$  und  $k=2$  (wofür in Abb. 2 ein Beispiel gegeben ist) lassen sich unter Berücksichtigung der Endzustände  $3,32224 \times 10^{14}$  Automaten generieren.

Leider war es nicht möglich den veröffentlichten Pseudocode zu implementieren, da es einige logische Fehler, die z. B. in Endlosschleifen resultierten, gab. Emails an die Autoren blieben unbeantwortet. [Be20] Daher wurde das Konzept des FIFA Generators übernommen und an den gegebenen Kontext angepasst.

### 3.2 Generierungsstrategie

Zwei Strategien der Generierung waren möglich: *Proaktiv* und *auf Abruf*. Die proaktive Strategie trennt die Generierung der Automaten von der Generierung der Aufgaben (Abb. 3). Es werden periodisch Automaten erzeugt und in einer Datenbank gespeichert, sodass bei einer Aufgabengenerierung in der Datenbank nach einem geeigneten Automaten gesucht werden kann. In dem Fall, dass kein geeigneter Automat gefunden wird, wird auf einen Fallback-Automaten zurückgegriffen. Dies sind die Automaten, die in den analysierten Aufgaben genutzt werden. Die „auf Abruf“-Strategie generiert nur dann einen Automaten, wenn auch eine Aufgabe generiert wird. Hier könnte durch einen internen Counter nach  $x$  Generierungen ein Abbruch und Zurückgreifen auf einen Fallback-Automaten realisiert werden. Die Herausforderung liegt dabei im Setzen des Schwellwerts  $x$ . Die Wartezeit pro Aufgabe ist davon abhängig, wie schnell ein geeigneter Automat gefunden wird, was bei einer zufälligen Generierung ohne Abbruch im Worst Case nie der Fall ist. Der Schwellwert für einen Abbruch darf aber nicht zu niedrig gesetzt werden, um seltener auf Fallback-Automaten zurückgreifen zu müssen.

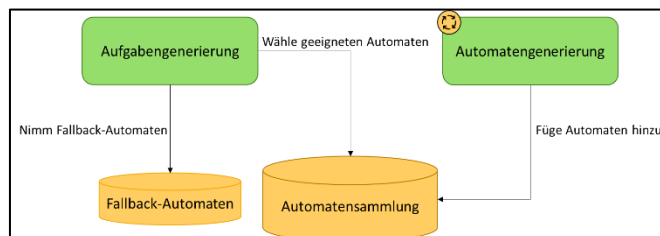


Abb. 3: Hinzufügen und Abrufen von Automaten bei der proaktiven Generierung

Durch das immer neue Generieren von Automaten in der „auf Abruf“-Strategie wird mit jedem generierten aber verworfenen Automaten Rechenleistung und Zeit verschwendet. Ein generierter Automat kann für die gewollte Aufgabe ungeeignet, für eine andere jedoch verwendbar sein.

Daraus folgend wurde sich für die proaktive Generierung entschieden. Ein kontinuierlich laufender Prozess erzeugt Automaten, bestimmt deren Eigenschaften und Zahlenwerte und speichert sie in einer Datenbank. Bei der Aufgabengenerierung werden die vorhandenen Automaten nach ihren Attributen gemäß der Eignungsklassifizierung des Aufgabentyps selektiert und zufällig einer zurückgegeben.

### 3.3 Filterung ungeeigneter Automaten

Um keine Automaten zu speichern, die für keinen der implementierten Aufgabentypen geeignet sind, wird eine Heuristik genutzt, die Automaten aussortiert, bei denen mindestens ein Zahlenwert außerhalb des Intervalls über die Zahlenwerte aller Aufgabentypen liegt. Tab. 4 gibt die Intervalle aller Zahlenwerte an, wobei für die Grenzen von Übergangsdichte und relativer Finalzustände zwischen deterministischen und nichtdeterministischen Automaten unterschieden wird.

Zahlenwert	Deterministisch	Minimum	Maximum
n	Keine Unterscheidung	2	8
k	Keine Unterscheidung	2	3
Übergangsdichte	Deterministisch	0,142	0,67
Übergangsdichte	Nichtdeterministisch	0,125	0,67
Finalzustände (relativ)	Deterministisch	12,5%	50%
Finalzustände (relativ)	Nichtdeterministisch	10,9%	26%

Tab. 4: Minimum und Maximum der Zahlenwerte über alle Aufgabentypen

Diese simple Filterung sortiert nicht alle ungeeigneten Automaten aus, doch würde die Überprüfung komplexerer Kriterien eine höhere Rechenzeit pro Automat bedeuten. Mit der beschriebenen Heuristik wurde ein guter Mittelwert zwischen Filterung und Einfachheit gefunden. In der Implementierung wird die Heuristik als eigenständiges Modul eingebunden, sodass sie durch eine andere ersetzt werden kann.

## 4 Der Webservice

Der *Automata Task Random Generator (AuTaRG)* wurde als Webtool umgesetzt. Diese Umsetzung erlaubt einen schnellen, orts- und betriebssystemunabhängigen Zugriff ohne die Notwendigkeit einer Installation oder dem Zugang zu einem bestimmten PC. Die Architektur ist in Abb. 4 gegeben und wird im Folgenden näher erläutert.

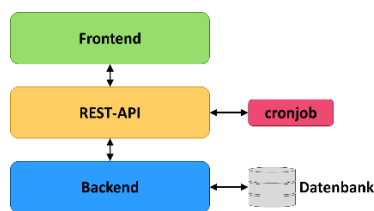


Abb. 4: Die Architektur des Webtools

Das Backend ist in Python geschrieben und setzt sich aus mehreren Modulen zusammen. Einzelne Module sind für die Generierung von Automaten oder Aufgaben zuständig, erstellen Export-Dateien oder dienen als Kommunikationsschnittstelle mit der Datenbank. Dieser Aufbau erlaubt einfache Anpassungen, Wartungen und Erweiterungen.

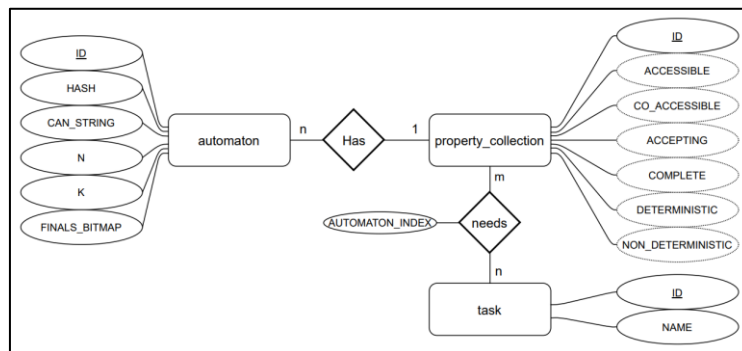


Abb. 5: Das Datenbankmodell

Abb. 5 zeigt das genutzte Datenbankmodell. Die Aufgabentypen und jede Kombination von Eigenschaften, die für einen Aufgabentyp benötigt werden, sind hinterlegt. Der `AUTOMATON_INDEX` wird zur Identifizierung der Automaten pro Aufgabe genutzt, da manche Aufgabentypen mehr als einen Automaten, ggf. mit verschiedenen Eigenschaften, benötigen. Nach der Generierung werden die Automaten auf ihre Eigenschaften untersucht und bei der Speicherung in der Datenbank mit der ID der dazugehörigen Eigenschaftenkombination verknüpft. Wird eine Aufgabe generiert, werden die entsprechenden Automaten mit den geforderten Eigenschaften selektiert und zufällig ein Automat zurückgegeben. Die so erstellte Aufgabe wird anschließend auf ihre Eignung für den Aufgabentyp geprüft.



Das Frontend ist als Single-Page-Application umgesetzt, welche Anfragen zur Generierung von Aufgaben oder Automaten durch die REST-API an das Backend sendet und die empfangenen Daten für den Nutzer aufbereitet und darstellt. Diese REST-API wird auch vom cronjob genutzt, der durch diese Schnittstelle einmal stündlich die Generierung von 100 kanonischen Strings mit zufälligen  $n$  und  $k$  im Backend anstößt. Diese werden zunächst ohne Endzustände erstellt und anschließend alle Endzustandskombinationen auf ihre Eignung getestet. Es resultierten durchschnittlich 34 geeignete Automaten aus einem kanonischen String. Die 100 Generierungen wurden durch Kalibrierung ermittelt, da sie eine gute Balance zwischen stündlichem Rechenaufwand und generierten, geeigneten Automaten bieten (durchschnittlich 3361,27 Automaten in 5 Sekunden).

Zusätzlich kann die Automaten generierung durch das Frontend manuell, mit der Möglichkeit konkrete Parameter für die Zahlenwerte vorzugeben, angestoßen werden. Dies kann genutzt werden, wenn AuTaRG neu aufgesetzt wird oder viele Automaten für einen bestimmten Aufgabentyp gebraucht werden. Auch das Generieren eines einzelnen Automaten ist möglich. Dazu werden im Frontend, neben dem Transitionsdiagramm und dem kanonischen String, die Eigenschaften des Automaten aufgelistet. Weiterhin kann die Akzeptanz von Worten getestet werden.

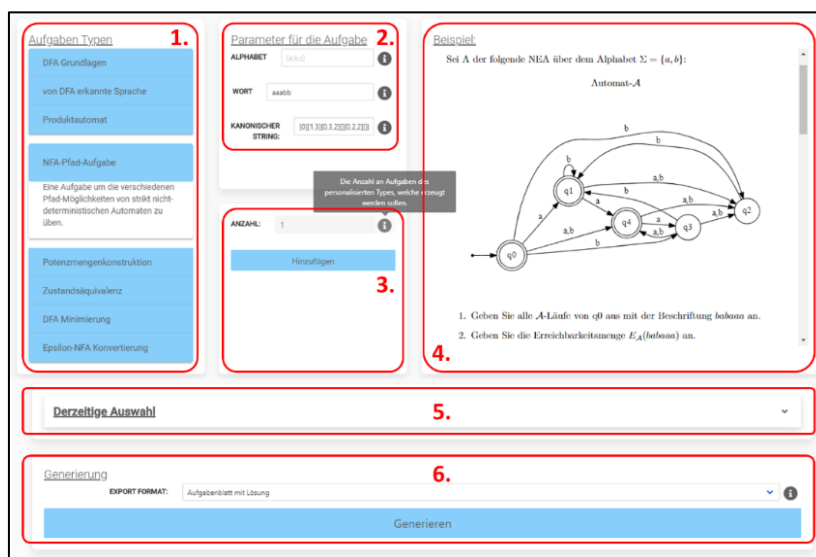


Abb. 6: Das Interface zur Aufgabengenerierung

Das sechsteilige Interface zur Aufgabengenerierung ist in Abb. 6 dargestellt. Zu Beginn wird im ersten Teil ein Aufgabentyp ausgewählt, für den Aufgaben generiert werden sollen. Soll eine Aufgabe mit einem bestimmten Automaten generiert werden, können im zweiten Teil der kanonische String eingefügt und ein Alphabet und weitere Parameter für die Aufgabenkonfiguration angegeben werden. Sollen stattdessen eine oder mehrere

Aufgaben mit zufälligen Automaten generiert werden, wird die gewünschte Anzahl im dritten Teil angegeben. Mit einem Klick auf den *Hinzufügen*-Button werden die angegebenen Parameter validiert, ehe die gewünschten Aufgaben der Auswahl angehängen werden.

Der vierte Teil zeigt eine statische Beispielaufgabe des ausgewählten Typs, während im fünften Teil die Auswahl eingesehen und einzelne Aufgaben wieder entfernt werden können. Abschließend wird im sechsten Teil das gewünschte Exportformat (PDF, LaTeX-Code oder Moodle-XML) ausgewählt und mit dem Klicken des *Generieren*-Buttons eine Anfrage an das Backend geschickt, die als Antwort eine Datei im gewünschten Format mit den geforderten Aufgaben zurückgibt, deren Download vom Frontend angestoßen wird.

Für den Export als PDF kann angegeben werden, ob die Lösung mitgegeben werden soll oder nicht, während im LaTeX-Code eine Variable eingesetzt wird, die festlegt, ob die Musterlösung beim Kompilieren generiert wird. Der Export als Moodle-XML ist gegeben, um die Aufgaben in Moodle-Quizze zu importieren. Zur Anpassung an verschiedene Fragetypen kann das genaue Format der XML im Backend definiert werden.

## 5 Performance und Nutzerevaluation

Die Performance der Automaten generierung wurde durch 100 Generierungen von 100 Automaten bei einer leeren Datenbank mit und ohne Nutzung der Heuristik zur Filterung getestet. Abb. 7 gibt die durchschnittlich benötigte Zeit zur Generierung eines Automaten in Millisekunden für beide Testläufe an.

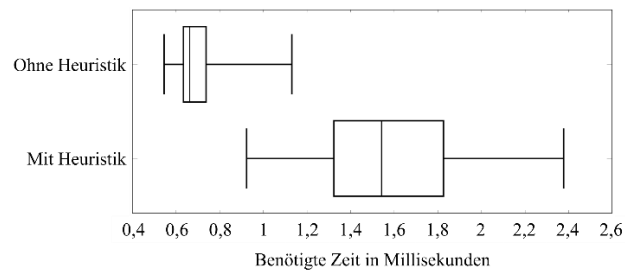


Abb. 7: Durchschnittlich benötigte Zeit zur Generierung eines Automaten mit und ohne Heuristik

Die Vorteile der Filterung ungeeigneter Automaten durch die Heuristik (weniger Suchdurchläufe und geringere Speicherkomplexität) rechtfertigen die geringfügig längere Laufzeit bei der Generierung.

Die Aufgabengenerierung wurde für je 500 Aufgaben zu jedem Aufgabentyp in drei Einzelschritten gemessen: Die Generierung der Aufgaben, das Rendern der Automaten als

Transitionsdiagramm und die Exporte als *LaTeX-Code*, *ein PDF pro Aufgabe* und *ein PDF für alle Aufgaben*. Der Export als Moodle-XML wurde nicht gemessen, da dieser nur für eine Aufgabe definiert ist (mehr dazu im Ausblick). Tab. 5 gibt die Ergebnisse wieder.

Aufgabentyp	Benötigte Zeit in Sekunden					$\Sigma$
	Aufgaben- generierung	Rendering	Exporte			
			LaTeX- Code	Einzelnes PDF	Separate PDFs	
<i>Einführung</i>	19,98	28,41	0,70	3,89	6,33	59,31
<i>Spracherkenn.</i>	7,44	28,71	0,28	2,22	4,26	42,91
<i>Produktauto.</i>	36,57	78,43	0,80	4,97	8,14	128,91
<i>NFA-Pfade</i>	183,78	30,81	0,44	3,01	5,12	223,16
<i>Potenzmenge.</i>	392,86	60,48	0,61	3,80	5,84	463,59
<i>Zustandsäqui.</i>	12,33	18,02	1,25	3,19	5,10	39,89
<i>Blockverfein.</i>	10,33	92,94	1,28	6,38	8,81	119,74
<i><math>\epsilon</math>-NFA</i>	173,38	56,97	0,48	3,81	6,12	240,76

Tab. 5: Benötigte Zeit für 500 Aufgaben pro Aufgabentyp in den Schritten Generierung, Rendering und den drei Exporten

Die Komplexität der Aufgabenkriterien (siehe 2.3) und die automatische Berechnung der Lösung hat einen starken Einfluss auf die benötigte Zeit der Aufgabengenerierung. Die längste Generierung (500 Aufgaben zur *Potenzmengenkonstruktion*) dauert etwas weniger als acht Minuten. Das händische Erstellen, Lösen und Übersetzen in LaTeX-Code einer einzigen Aufgabe dieses Aufgabentyps dauert erfahrungsgemäß länger. Die Nutzung von AuTaRG erweist sich somit als große Zeiteinsparung.

Das Frontend von AuTaRG wurde von sechs ehemaligen und zukünftigen FoSAP-Assistenten mit der Thinking-Aloud-Methode und einem abschließenden fragebogen-gestützten Interview evaluiert. Das Frontend wurde als verständlich und gut nutzbar bewertet. Dabei wurde die Anzeige einer Beispielaufgabe, obwohl sie statisch ist, als besonders hilfreich hervorgehoben. Anpassungen zur Beseitigung von Darstellungsfehlern wurden vorgenommen und als Reaktion auf weitere Wünsche der Assistenten an das Tool die Funktionalität zur grafischen Erstellung eines einzelnen Automaten hinzugefügt. Weiterhin wurde die Korrektheit der generierten Aufgaben samt Musterlösung geprüft und bestätigt.

AuTaRG wurde als sehr nützlich für den Übungsbetrieb empfunden. Die schnelle Generierung von zufälligen Aufgaben entlastet die Assistenten darin, sich keine neuen Aufgaben ausdenken zu müssen. Das Einsparen von Zeit durch die schnelle Generierung wurde bestätigt. Die implementierten Exportformate sind gut gewählt und umgesetzt. Für Sprechstunden können schnell nach Bedarf Aufgaben und als Anschauungsbeispiele für z. B. die Vorlesung einzelne Automaten generiert werden. Ein Teilnehmer würde beim Export von LaTeX-Code zwei Dateien, eine mit Lösung, eine ohne, bevorzugen, da sich dies besser in das Template des Lehrstuhls integrieren lässt. Dies kann aktuell durch das Kopieren des Lösungs-codes in eine neue Datei umgesetzt werden.

## 6 Fazit und Ausblick

Dieser Beitrag zeigt, wie Automaten und automatenbasierte Aufgaben zufällig automatisiert generiert werden können. Am Beispiel der Aufgaben der Vorlesung Formale Systeme, Automaten, Prozesse (FoSAP) wurden Aufgabentypen und deren Lernziele identifiziert. Anhand von Eigenschaften und Zahlenwerten wurde präsentiert, wie für jeden Aufgabentyp Eignungsklassifikationen für Automaten erstellt wurden. Zur zufälligen Generierung von Automaten wurden verschiedene Generatoren aufgearbeitet und angelehnt an den FIFA Generator ein eigenes Verfahren entwickelt. Dieses Verfahren wurde im Webtool *Automata Task Random Generator (AuTaRG)* implementiert und eine proaktive Generierung der Automaten umgesetzt, sodass bei der Aufgabengenerierung aus einem großen Pool zufällig ein Automat ausgewählt wird. AuTaRG wurde in der Nutzerevaluation mit FoSAP-Assistenten als sehr nützlich für den Übungsbetrieb und darüber hinaus empfunden. Die schnelle Generierung von zufälligen Aufgaben entlastet sie darin, sich keine neuen Aufgaben ausdenken zu müssen und spart Zeit ein.

Es ist zunächst nicht geplant, dass Tool Studierenden zur Verfügung zu stellen. Stattdessen werden Questiontypes für Moodle entwickelt, welche die Aufgabentypen als Fragen für Moodle-Quizze mit einer automatisierten Bewertung umsetzen. Während Fragen zur Wortakzeptanz als Multiple-Choice Frage umgesetzt werden, bedarf es bei anderen Aufgaben einer anderen Eingabe z. B. einer grafischen, was in Moodle aktuell nicht möglich ist. Auch das Hinzufügen weiterer Exportformate für die Verwendung in anderen Lernmanagement Systemen ist möglich.

Ein weiteres interessantes Kriterium für die generierten Aufgaben ist die Schwierigkeit, basierend auf den verwendeten Automaten. Dazu können z. B. die Antworten auf die Aufgaben ausgewertet werden. Eine erste Verwendung von AuTaRG im Übungsbetrieb der FoSAP Vorlesung findet im Sommersemester 2020 statt.

## Literaturverzeichnis

- [Be20] Bergerbusch, T.: Generating Computer Science Exercises with Suitable Automata, Master Thesis, 2020.
- [Ch04] Champarnaud, J. et al.: Random Generation Models for NFAs. *Journal of Automata, Languages and Combinatorics*. 9, 203-216, 2004.
- [Fe18] Ferreira, M.; Moreira, N.; Reis, R.: Forward Injective Finite Automata: Exact and Random Generation of Nonisomorphic NFAs. In: Konstantinidis S., Pighizzini G. (eds) *Descriptive Complexity of Formal Systems. DCFS 2018. Lecture Notes in Computer Science*, vol 10952. Springer, Cham, 2018.
- [Ku20] Kurdi, G. et al.: A Systematic Review of Automatic Question Generation for Educational Purposes. *International Journal of Artificial Intelligence in Education* 30, 121–204, 2020.
- [Le95] Leslie, T.: *Efficient Approaches to Subset Construction*, 1995.

## Self-Assessment mit High-Information Feedback<sup>1</sup>

Joerg M. Haake<sup>2</sup>, Niels Seidel<sup>2</sup>, Heike Karolyi<sup>3</sup>, Lihong Ma<sup>2</sup>

**Abstract** Die Gestaltung und Umsetzung von skalierbaren lernförderlichen Feedbacks gilt als große Herausforderung im digitalisierten Hochschulstudium: High-Information Feedback (HIF) kann als didaktisches Element zur Förderung der Selbsteinschätzung und Steigerung des Lernerfolgs beitragen. Self-Assessment mit HIF soll Lernenden in einem iterativen Prozess ermöglichen den erreichten Lernstand besser einzuschätzen und selbst die Rolle des Feedback-Gebenden zu übernehmen.

**Keywords:** online Self-Assessment, High-Information Feedback, selbstreguliertes Lernen

### 1 Einleitung

Feedback ist ein wirksames Instrument, um selbstreguliertes Lernen und die Selbstbestimmungstheorie des Lernens [DR93] zu unterstützen [WHZ20]. Eine gute Feedback-Qualität für jeden Lernenden sicherzustellen ist aufwändig und mindert daher die Skalierbarkeit, die in großen Lehrveranstaltungen unverzichtbar ist. Ebenso ist Wiederholbarkeit wichtig für das Üben und inkrementellen Kompetenzaufbau, wobei jede Wiederholung angepasstes Feedback erfordert und damit die Skalierbarkeit erschwert.

Dieser Beitrag veranschaulicht die Entwicklung eines skalierbaren Ansatzes für wiederholbare Übungen von komplexen Aufgaben zu Reorganisations- und Transferleistungen oder Reflexion bzw. Problemlösen mit lernförderlichem unmittelbarem Feedback in digitalen Lernumgebungen.

Gängige Ansätze zur Bereitstellung von Feedback für Lernende nutzen als Feedback-Gebende entweder Menschen (Dozenten/Tutoren oder Peers) oder Software (z.B. maschinell korrigierbare Aufgaben wie Quizzes/MC-Tests, parametrisierbare Aufgaben oder Rückmeldebaum- und Pfad-basierte Ansätze). Als neuen Lösungsansatz stellen wir in diesem Beitrag eine Kombination von Mensch und Software vor. Statt Dozenten/Tutoren oder Peers sollen Lernende ihre eigene Leistung bewerten (Self-Assessment), was die Skalierbarkeit, unmittelbares Feedback und Wiederholbarkeit unterstützt. Dabei werden sie von der Lernumgebung bei der Durchführung eines Self-Assessments mit passendem Feedback unterstützt.

---

<sup>1</sup> Diese Forschungsarbeiten wurden durch den Forschungsschwerpunkt "Digitalisierung, Diversität und Lebenslanges Lernen - Konsequenzen für die Hochschulbildung" (D<sup>2</sup>L<sup>2</sup>) der FernUniversität in Hagen unterstützt.

<sup>2</sup> FernUniversität in Hagen, Fakultät für Mathematik und Informatik, Universitätsstr. 11, 58084 Hagen, {joerg.haake, niels.seidel, lihong.ma}@fernuni-hagen.de

<sup>3</sup> FernUniversität in Hagen, Kultur- und Sozialwissenschaften, Universitätsstr. 33, 58084 Hagen, heike.karolyi@fernuni-hagen.de

## 2 Self-Assessment mit High Information Feedback

Self-Assessments (SA) in Online-Lernumgebungen bieten Lernenden vielfältige Übungsmöglichkeiten, um sich mit Lernzielen, Anforderungen und Bewertungskriterien auseinanderzusetzen und Lernstand sowie Leistung besser einschätzen zu lernen [Har17]. Im Rahmen eines Self-Assessments (SA), also einer sich selbst bewertenden Rückmeldung des Lernenden an sich selbst im Sinne einer angeleiteten Reflexion [IJ10, S. 244], übernehmen Lernende selbst die Rolle des Feedback-Gebers und entwickeln anhand vorgegebener Bewertungskriterien ein Verständnis davon, was „eine gute Aufgabebearbeitung im Fach [...] auszeichnet“ [Har17, S. 204]. Gerade in Online-Kursen ist es entscheidend „Feedback als festes Element in das didaktische Design zu integrieren.“ [Har17, S. 211]. Ramaprasad definiert: „Feedback is information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way“ [Ram83, S. 4]. Feedback kann sowohl den Lernfortschritt, als auch die Selbstregulationsfähigkeiten der Lernenden unterstützen und damit zu nachhaltigem Lernerfolg führen. Wisniewski et al. stellen dabei wiederholt fest, dass Belohnung oder Strafe sich negativ auf die intrinsische Motivation auswirken, indem Selbstregulationsstrategien und Selbstmotivation gemindert werden [DR93]. High-Information Feedback (HIF) [WZH20, S.12] hingegen hat eine lernförderliche Wirkung auf den Lernprozess. Durch Informationen zu Aufgabe, Lösungsprozess und (manchmal) zum Niveau der Selbstregulierung profitieren Studierende stark von einem HIF, wenn es ihnen hilft zu verstehen, welche Fehler sie warum gemacht haben und was sie tun können, um sie in Zukunft zu vermeiden [HT07, WZH20]. Ein effektives Feedback adressiert verschiedene Perspektiven, um die Erweiterung von Wissen und das Verständnis von Sachverhalten differenziert zu betrachten [HT07]. Feed Up konzentriert sich auf das Verhältnis von Lernstand und Lernziel. Feed Back veranschaulicht den Lernzuwachs, indem es den letzten Lern- und Entwicklungsstand mit dem aktuellen vergleicht und den Zielstatus erklärt. Feed Forward adressiert die Adaption weiterführender Zielsetzungen, es unterstützt Selbstregulation über den Lernprozess hinweg und liefert ein detailliertes Verständnis darüber, was verstanden wurde und was nicht. Um ein Feedback für Lernende nutzbar zu machen, ist es in einen Feedback-Prozess einzubinden [Car13], es unterstützt den Lernerfolg nachweislich, wenn es die Selbstregulationsfähigkeit stärkt und dem Lernenden verdeutlicht, was eine gute Leistung ausmacht [HT07, WZH20, Har17].

Durch die Kombination von SA und HIF in Form einer SAMHIF-Aufgabe üben Lernende die konstruktive Lösung von komplexen Aufgaben mit Wiederholungsmöglichkeiten und erhalten differenzierte Rückmeldungen zu typischen Fehlerquellen, die als Lösungsbausteine beschrieben sind. Dabei werden Feedbacks so gestaltet, dass neben Informationen zur strukturierten und systematisierten Bearbeitung der Aufgabe und zum Nachvollziehen des Lernprozesses auch die Selbstregulierung durch Tipps zu Lernstrategien befördert wird.

Dazu verläuft die Bearbeitung in einem iterativen Prozess (siehe Abb. 1). Zuerst zeigt das Übungssystem die Aufgabenstellung (und damit implizit die Lernziele) an und fordert die

Lernende zur Erstellung einer Lösung auf. Die Lernenden erstellen nun eine Lösung, z.B. auf Papier. Anschließend zeigt das Übungssystem die Bewertungskriterien an und fordert die Lernende auf, ihre Lösung anhand verschiedener Kriterien zu bewerten. Die Kriterien modellieren Teilziele, deren Wahrnehmung und Einschätzung bzgl. des Erfüllungsgrads als Feedback des erreichten Verständnisses dient. Nachdem Lernende ihre Selbstbewertung eingegeben haben erzeugt das Übungssystem ein HIF als passende Rückmeldung. Wenn die Lernenden mindestens ein Kriterium einer korrekten Lösung nicht erfüllt haben regt das Übungssystem zur Überarbeitung an. Entscheidet sich die Lernenden zur Überarbeitung der Lösung wird die nächste Iteration der Aufgabebearbeitung eingeleitet. Hier realisiert die Anzeige von aktueller Selbstbewertung und HIF eine Feed Up Information zur Identifikation noch nicht erreichter Lernziele.

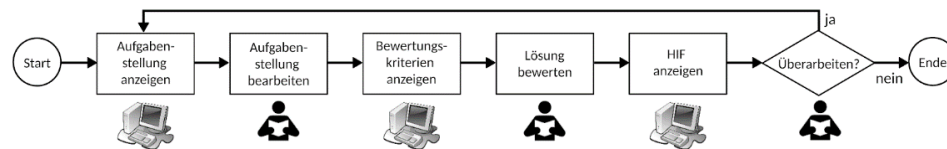


Abb. 1: Iterative Bearbeitung einer SAMHIF-Aufgabe in fünf Phasen

### 3 Verwandte Arbeiten

Gängige Methoden erzeugen Feedback entweder durch Menschen oder durch den Computer. Der Vorteil des von Menschen erzeugten Feedbacks liegt in der erreichbaren Qualität. So können Kursbetreuende bei der Korrektur von Übungsaufgaben für das Individuum zugeschnittenes HIF mit geringem Aufwand erzeugen. Aufgrund der beschränkten Anzahl von Betreuenden skaliert dies aber nicht, führt zu langen Wartezeiten bis ein Feedback erfolgt und hemmt so in der Praxis eine wiederholte Überarbeitung mit neuem Feedback. Um den Betreuungseinsatz zu umgehen können in bestimmten didaktischen Szenarien Peer-Review-Verfahren eingesetzt werden. Allerdings ist die Erstellung der Review-Instruktionen und das Bereitstellen von Bewertungshinweisen für Reviewer aufwändig. Zudem sind bei der Erstellung von Reviews durch Peers, falsche oder unpassende Feedbacks nicht auszuschließen und erfordern einen zusätzlichen Qualitätssicherungsaufwand.

Maschinell korrigierbare Aufgabentypen (z.B. MC oder Programmieraufgaben) bieten eine sofortige Bewertung, diese sind allerdings nur mit eingeschränkten Eingabeoptionen (Checkboxen oder Programmzeilen) mit darauf bezogenen, oft wenig konstruktiven Bewertungen und Hinweisen möglich. Eine Überarbeitung führt bei variierenden Antwortauswahlen ggf. zu neuen Bewertungshinweisen. Neuere Entwicklungen gehen hier einen Schritt weiter. Übungsaufgaben in STACK [WPH19] definieren einen Fragetext, Variablen, einen Rückmeldebaum für verschiedene Fälle und dazu passendes, situatives Feedback sowie die Gegenüberstellung der Eingabe mit der Musterlösung. Aufgabenvarianten erlauben Wiederholungen, ohne eine identische Aufgabe zu präsentieren. Pfadbasierte Aufgaben [SS+15] unterstützen die Anpassung einer konkreten

Aufgabenbearbeitung in Abhängigkeit vom Erfolg der bisherigen Bearbeitungsschritte und ermöglichen situatives Feedback. Die möglichen Schritte entsprechen Teilaufgaben, die durch Bedingungen miteinander verknüpft sind. Die Nutzung von Parametern erlaubt die Erzeugung individueller Aufgabenvarianten und damit auch wiederholtes Üben. STACK und pfadbasierte Aufgaben teilen den hohen Erstellungs- und Testaufwand sowie die Notwendigkeit, dass Lernende die Aufgabe in der computergestützten Lernumgebung lösen müssen.

Der von uns vorgeschlagene Aufgabentyp SAMHIF stellt eine hybride Lösung dar, bei der Lernende das HIF im Rahmen eines Self-Assessments nutzt. Dabei unterstützt das Übungssystem ein Selbst-Feedback in einem inkrementellen Lernprozess, in dem Lernende die Aufgabe auch außerhalb des Übungssystems lösen können. Dieses Vorgehen erlaubt beliebige Skalierung mit potenziell hoher Qualität. Der Aufwand der interdisziplinären Erstellung von SAMHIF wird im Folgenden beschrieben.

## 4 Umsetzung

Die Online-Lehre gilt aus bildungsorganisatorischer Perspektive als äußerst effektiv; ein einmal entwickelter Kurs kann in der Regel wiederholt angeboten und mit wenig Aufwand aktualisiert werden. Während bei der Erstellung von Texten, Videos oder Tests ein einmaliger Aufwand entsteht, sind Betreuungsaufgaben und Ressourcen für die Generierung differenzierter Feedbacks nur schwer zu skalieren [Har17]. Dies führte zu Überlegungen, wie man wirkungsvolle Selbst-Assessments entwickeln kann, die den Lernerfolg im Hochschulstudium nachhaltig steigern, indem Lernende sich sowohl in den Kenntnissen zur Domäne, als auch in Bezug zum eigenen Lernverhalten üben. Ausgangspunkt waren domänenspezifische Aufgaben aus einem Kurs der ersten Studienphase, die bislang als Selbstlernaufgaben in Kombination mit Musterlösungen angeboten wurden. Diese zeichnen sich durch Operatoren der zweiten oder dritten Stufe aus, die Reorganisations- und Transferleistungen oder Reflexion bzw. Problemlösen fordern. Lernende erstellen dabei Lösungen mit komplexen sowie ausführlichen Beschreibungen, die einen hohen Korrekturaufwand aufweisen. Im Zuge der Digitalisierung von zusätzlichem Übungsmaterial sollten diese Aufgaben neu gestaltet und für eine bessere Unterstützung des selbstregulierten Lernens im Hochschulstudium mit skalierbaren Feedbackmöglichkeiten nutzbar gemacht werden. Die Entwicklung der skalierbaren Feedbacks wird in mehreren Phasen umgesetzt, die durch die hier diskutierte erste Evaluation begleitet wird.

Die vorgestellten Self-Assessment wurde im Rahmen des Pflichtkurses »Betriebssysteme und Rechnernetze« des Fernstudiengangs B.Sc. Informatik im Wintersemester 2019/2020 mit dem Standard-Plugin Test (mod\_quiz) in Moodle implementiert und eingesetzt. In einem ersten Schritt wurden die Aufgabenstellungen in das Plugin übertragen. Da bei SAMHIF die Kontrolle der Lösung den Studierenden obliegt, werden die Lösungen nicht in einem digitalisierten Format eingefordert. Von größerer Relevanz ist in der ersten Entwicklungsphase die Ausarbeitung von Bewertungskriterien, die häufige Fehlerquellen der



Lernenden direkt adressieren. Die Bewertungskriterien waren dabei schon bei der Aufgabenbearbeitung sichtbar und das HIF wurde auf ein Lernzielfeedback beschränkt.

Von den 281 aktiven der insgesamt 305 Kursteilnehmenden haben 190 Personen mindestens eine von insgesamt 43 Aufgaben dieses Typs bearbeitet. 66 der 190 Studierende beteiligten sich an einer Online Befragung [SKH20] zur Bewertung der Qualität der Self-Assessment mit Lernzielfeedback. Laut der Moodle-Logdaten haben 119 Studierende mindestens eine Aufgabe mehrfach bearbeitet. Als Begründung für die mehrfache Wiederholung gaben die Studierenden in der Befragung an, Defizite ausgleichen zu wollen (65%), eine richtige Lösung erbringen zu wollen (58%), Wissen aufzufrischen (54%) und sich auf die Prüfung vorzubereiten (42%). 29% der Befragten gaben an Aufgaben generell mehrfach zu bearbeiten und 15% hatten die Wiederholung in ihrem individuellen Lernplan vorgesehen. Die meisten Wiederholungen erfolgten innerhalb weniger Minuten, wobei sich die Selbsteinschätzung der eigenen Lösung in den nachfolgenden Aufgabenbearbeitungen um 40,1 % verbesserte. Die Formulierung von Aufgabenstellung, Arbeitsanweisung, Lösungseigenschaften sowie des Lernzielfeedbacks bewerteten die Befragten jeweils mit über 90% als verständlich. Die Darstellung der Lösungseigenschaften wurde von 82 % der Befragten als hilfreich für die Aufgabenbearbeitung empfunden. Es wurde angeregt, die Bewertungskriterien durch Distraktoren zu ergänzen und erst nach der Lösungseinreichung einzublenden. Das Lernzielfeedback wurde von den Befragten insgesamt oft als hilfreich angesehen, um Fehler in der eigenen Lösung schnell zu identifizieren und zu beheben. 22 % verwiesen darauf, Fehler auch ohne das Lernzielfeedback finden zu können. Je nach Art des Fehlers und der verfügbaren Lernmaterialien diene das Feedback als ergänzende Unterstützung.

Die Ergebnisse der Befragung unterstreichen die anhand der Nutzungszahlen belegte Akzeptanz durch eine qualitative Bewertung der SAs. Alle Bestandteile des Aufgabentyps wurden positiv bewertet und lassen darauf schließen, dass die SAMHIF-Aufgaben die Lernenden zur Fehlersuche, -korrektur und Lösungs-Überarbeitung anregen.

## 5 Zusammenfassung und Ausblick

Der in diesem Beitrag vorgestellte Aufgabentyp ist mit wenig Aufwand zu erstellen, skalierbar einsetzbar, ermöglicht eine effektive Bewertung der eigenen Lösung durch die Lernenden und bietet reichhaltiges Feedback, welches ein detailliertes Verständnis zum eigenen Lernstand vermittelt. Aus den Erfahrungen der ersten Entwicklungs- und Erprobungsphase wird SAMHIF weiterentwickelt und für eine weitere Felduntersuchung vorbereitet. Die Evaluation zeigt, dass die Durchführung der fünf Phasen im Bearbeitungsprozesses eine klare visuelle Trennung der SA Phasen im Test erfordert, um die domänenspezifische Aufgabenbearbeitung und die Präsentation der Feedbacks besser zu separieren. Damit wird eine weitere Interaktionsebene zur Reflexion der Leistung durch die Lernenden anhand der HIFs realisiert. Die Präsentation der Lösungskriterien führte zu einer konfirmatorischen Bewertung durch Anklicken erfüllter Lösungseigenschaften. In

der Weiterentwicklung werden diese Lösungskriterien nun mit weiteren Feedbackinformationen zum Lernprozess und zu passenden Lernstrategien angereichert, sodass die Lernenden sowohl in Bezug zur Domäne als auch zum Selbstregulierten Lernen unterstützt werden können. Die Entwicklung des Feedback Content wurde zunehmend systematisiert, sodass die dreiteiligen HIFs (Informationen zu Lernziel, Lernprozess und Lernstrategie) effektiv erstellt werden können. Informationen zu Lernstrategien und Lernprozess können als sich wiederholende Textbausteine aus einer Auswahl ergänzt werden.

Differenzierte Feedbacks werden im Moodle Test nur bei Anklicken einer Auswahlmöglichkeit präsentiert. Lernende durchschauen dies und klicken nicht wahrheitsgemäß an, um mehr Informationen zu erhalten. Ein verbessertes Test-Plugin soll dies verhindern.

## Literaturverzeichnis

- [Car13] Carless, D.: Sustainable feedback and the development of student self-evaluative capacities. In S. Merry, M. et.al. (Hrsg.), *Reconceptualising feedback in higher education. Developing dialogue with students*, 113–122. Routledge, London, 2013.
- [DR93] Deci, E. L.; Ryan, R. M.: Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39(2), 223-238, 1993.
- [Har17] Hartung, S.: Lernförderliches Feedback in der Online-Lehre gestalten. In H. R. Griesehop & E. Bauer (Hrsg.), *Lehren und Lernen online*. Springer, Wiesbaden, 199–217, 2017.
- [HT07] Hattie, J.; Timperley, H.: The Power of Feedback. *Review of Educational Research*, 77(1), 81–112, 2007.
- [IJ10] Ibabe, I.; Jauregizar, J.: Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59, 243–258, 2010.
- [Ram83] Ramaprasad, A.: On the definition of feedback. *Systems Research and Behavioral Science*, 28(1), 4–13, 1983.
- [SKH20] Seidel, N.; Karolyi, H.; Haake, J.M.: Evaluation einer Implementierung von Self-Assessments mit Lernziel-Feedback in Moodle. *Research Report 1/2020*. FernUniversität in Hagen, 2020.
- [SS+15] Schwinning, N. et.al.: Aufwand und Nutzen parametrisierbarer, pfadbasierter Aufgaben. In: Pongratz, H.; Keil, R. (Hrsg.), *DELFI 2015*. Bonn: GI e.V., 325-327, 2015.
- [WPH19] Weigel, M.; Podgayetskaya, T.; Hübl, R.: Digitalisiertes Feedback für verschiedene Frageformate von Online-Übungsaufgaben. In: Pinkwart, N.; Konert, J. (Hrsg.), *DELFI 2019*. Bonn: GI e.V., 121-126, 2019.
- [WZH20] Wisniewski, B.; Zierer, K.; Hattie, J.: The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology*, 10, 2020.

## Umsetzung digitaler Prüfungen mit bwLehrpool an der Universität Freiburg

Dirk von Suchodoletz<sup>1</sup>, Sven Slotosch<sup>1</sup>, Christian Rößler<sup>1</sup> und Steffen Ritter<sup>2</sup>

**Abstract:** bwLehrpool entwickelt sich aus einem baden-württembergischen Landesprojekt zum Betrieb von netzwerk-geboteten PC-Pools zu einer verstärkt eingesetzten Plattform für E-Assessments. Der landesweit angebotene Dienst hat die Nutzung von Rechnerpools mittels Desktop-Virtualisierung flexibilisiert und schafft so die infrastrukturellen Voraussetzungen, um digitale Prüfungsszenarien in größerem Maßstab umzusetzen. Finden Prüfungen zumeist über browserbasierte Formate statt, erweitert bwLehrpool den Werkzeugkasten, indem Prüfungen mit beliebigen, vorinstallierten Programmen abgehalten werden können. Ein Prüfungsmodus in bwLehrpool in enger Verzahnung mit einem Learning Management System (LMS) erlaubt weitergehende Absicherungen der Prüfungsumgebung. Poolräume lassen sich hierfür mittels Webschnittstelle binnen weniger Minuten flexibel in einen abgesicherten Modus versetzen. Dieser Erfahrungsbericht gibt einen Überblick über durchgeführte E-Prüfungen an der Uni Freiburg in den letzten Jahren und zeigt auf, welche zusätzlichen Varianten für die Umsetzung elektronischer Prüfungen bestehen.

**Keywords:** E-Prüfungen, bwLehrpool, PC-Pools, kooperativer Landesdienst

### 1 Motivation

E-Assessments werden zunehmend wichtiger Bestandteil heutiger Lehre, da sie zahlreiche Vorteile für Lehrende und Studierende bieten. Nach anfänglichem Aufwand lässt sich mit Erfahrung und auf Basis bereits angelegter Fragenpools der Vorbereitungsaufwand reduzieren. Weitere Zeitersparnis stellt sich bei der Korrektur ein, da viele Fragetypen automatisiert durch das Prüfungssystem kontrolliert werden können. Lehrende müssen nur noch einen Bruchteil der Klausur selbst korrigieren und lediglich Kontrollfunktion ausüben (vgl. [VS09], S. 4). Ein großer Vorteil elektronischer Prüfungen ist der mögliche Einsatz multimedialer Inhalte. Es können Filmsequenzen und Bewegungsabläufe, fachspezifische Bildaufnahmen oder Sprachsamples in die Prüfung eingebaut werden (vgl. [EL20]). Dadurch wird das Fragenspektrum nicht nur erweitert, sondern auch wesentlich praxisnäher.

Es ist nicht mehr zielgerichtet, IT-gestützte Kurse zu Programmiersprachen oder speziellen Tools mit einer Papierklausur zu beschließen. So sollte es Standard werden,

---

<sup>1</sup> Albert-Ludwigs-Universität Freiburg, Fahrenbergplatz, 79085 Freiburg im Breisgau,  
dirk.von.suchodoletz@rz.uni-freiburg.de, sven.slotosch@rz.uni-freiburg.de, christian.roessler@rz.uni-freiburg.de

<sup>2</sup> Hochschule Offenburg, Badstraße 24, 77652 Offenburg, steffen.ritter@hs-offenburg

Aufgaben zu stellen, die den Einsatz der vorher im Kurs benutzten Werkzeuge einschließen.

Diese Entwicklungen erfordern entsprechende Infrastrukturen, die optimalerweise auf Basis bereits etablierter Ressourcen wie PC-Pools, LMS und bestehender geeigneter Räume aufbauen können. Hierzu war das Freiburger Universitätsrechenzentrum im Landesprojekt bwEKlausuren [Gi17] gemeinsam mit der Hochschule Offenburg aktiv, welches von 2015 bis Mitte 2017 vorangetrieben wurde und in einen allen Hochschulen angebotenen Landesdienst mündete. Fünf Jahre nach dem ersten Bericht auf der DELFI 2016 [Ri16] wollen die Autoren einen aktuellen Überblick geben und die laufenden Aktivitäten zusammenfassen. Der Text referiert hierzu die Erfahrungen nach inzwischen mehreren Jahren E-Klausuren in verschiedenen Formen der Durchführung.

## **2 Die Basisplattform und der Prüfungsmodus**

Die Einsatzgebiete von bwLehrpool erstrecken sich vom Betrieb klassischer Linux- oder Windows-Umgebungen an beliebigen Standorten über die Bereitstellung flexibler Kiosk-Systeme in großer Zahl – zur Recherche beispielsweise in Bibliotheken, ad-hoc für Konferenzen – bis hin zur Auslieferung standardisierter Arbeitsplätze. Eine lokale Installation der Betriebssysteme auf Arbeitsplatzrechnern wird vermieden. Die Umgebung wird über Netzwerk geladen und in Form Virtueller Maschinen, die als Virtuelle Lehr- und Arbeitsumgebungen bezeichnet werden, bereitgestellt und ausgeführt. Dieser Ansatz reduziert den Wartungsaufwand erheblich, lässt eine beliebige Anzahl von Maschinen zu und erlaubt das gleichzeitige Angebot eines breiten Spektrums an Softwareumgebungen.

bwLehrpool verlangt keine exklusive Installation und kann parallel zu bestehenden lokalen Betriebssystemen gebootet werden. Auf diese Weise können zusätzlich PC-Pools für E-Prüfungen aktiviert werden, die sonst in anderen Betriebsmodi arbeiten. Hierzu können Poolräume mittels Webschnittstelle binnen weniger Minuten flexibel auf einen Prüfungsmodus umgestellt werden.

Der Prüfungsmodus erlaubt verschiedene Absicherungen z.B. in Bezug auf Netzwerk oder USB. Damit wird unterbunden, dass Studierende während der Klausur nach Antworten auf die Klausurfragen im Internet suchen oder unerlaubterweise auf USB-Sticks mitgebrachte Unterlagen einsehen können. Je nach Konfiguration kann anstelle von Homeverzeichnissen ein anderes Dateisystem eingehängt werden, auf dem alle Teilnehmenden Schreibrechte besitzen, aber nur die selbst angelegten Dateien und Verzeichnisse sehen und modifizieren können. Damit lassen sich Abgabeverzeichnisse realisieren, in dem Teilnehmende Ergebnisse zur späteren Korrektur ablegen. Aufgrund fehlender Leserechte werden Betrugsversuche unterbunden und Abschreiben verhindert. Lehrende erhalten auf diesem Dateisystem erweiterte Rechte und können alle von den Studierenden erstellten Dateien einsehen.

### 3 Nutzungsszenarien

Bei der Modernisierung der PC-Pool-gestützten Lehre besteht das Ziel, eine breite Unterstützung mit elektronischen Werkzeugen und geeigneten desktop-basierten Lehr- und Arbeitsumgebungen umzusetzen. Zusätzlich können aktuell vier unterschiedliche Prüfungsszenarien realisiert werden:

Die "klassische" **E-Klausur** wird auf dem Prüfungs-ILIAS erstellt und durchgeführt. Per bwLehrpool werden Prüfungsrechner in den Poolräumen mit dem Server verbunden. Als zusätzliche Sicherheitsvorkehrung regelt der Safe Exam Browser (SEB)<sup>3</sup> den Zugriff. Die eingesetzten Fragetypen können durch multimediale Elemente ergänzt werden.

Die **E-Klausur plus** funktioniert ähnlich wie die "klassische" E-Klausur. Ergänzend wird weitere Software wie RStudio, SPSS u.ä. freigeschaltet und in Prüfungen genutzt.

Bei einer **VM-Prüfung** kann die zur Ausbildung verwendete virtuelle Maschine ohne besondere Modifikationen zur Prüfung eingesetzt werden, z. B. für Prüfungen an Spezialsoftware, GIS-Programmen oder für Programmieraufgaben. Die Prüfung wird so im gewohnten Lern- und Arbeitsumfeld abgelegt.

Bei der **Prüfung auf externem Server** werden die Rechner mit einem externen Server verbunden, auf dem die Klausur stattfindet. Hier dient ein bwLehrpool-Image mit installiertem SEB als Basis, um eine abgesicherte Verbindung zum Prüfungsserver herzustellen und Zugriffe auf andere Netzressourcen zu verhindern.

#### 3.1 bwLehrpool im Prüfungsmodus im Zusammenspiel mit einem LMS

Viele elektronisch basierte Prüfungen finden mittels eines LMS wie ILIAS oder Moodle statt. Am Standort Freiburg wird als Prüfungsplattform eine separate ILIAS-Umgebung als unabhängiger und speziell abgesicherter Prüfungsserver betrieben. Mit seiner umfangreichen Test- und Assessment-Funktionalität und zahlreichen Fragetypen hat sich ILIAS als Prüfungssystem bundesweit bewährt und erfährt besonders in diesem Bereich eine Weiterentwicklung [DK19]. Durch die Installation als separates Prüfungs-ILIAS wird das System unabhängig von den Update- und Wartungszyklen des Lehr-ILIAS betrieben. Durch den Anschluss an das LDAP-Verzeichnis der Uni entfällt ein vorheriges Verschicken von Accounts und Passwörtern.

Eine Umgebung zum Einsatz LMS-basierter Prüfungen hat ein minimalistisches Setting, bei dem lediglich ein Browser zur Verfügung steht. Sichergestellt werden muss in diesem Kontext, Aufrufe von (Dritt-)Seiten zu blockieren, die nicht vom LMS-Server im Rahmen der Prüfung ausgeliefert werden. In Freiburg kommt zu diesem Zweck die Software SEB zum Einsatz. Mit diesem ist es unkompliziert möglich, eingesetzte bwLehrpool-Images komplett abzusperrern. Es gelingt ebenfalls, definierte Drittsoftware einzubinden, die zur

---

<sup>3</sup> <https://www.safexambrowser.com>

Lösung von Aufgaben benötigt wird oder gezielt Websites freizugeben, die auch in der beruflichen Praxis genutzt werden.

### **Große Assessments**

Prüfungen mit über 100 Studierenden stellen auch das Prüfungssystem der Uni Freiburg vor größere Herausforderungen, da in der Regel über mehrere Räume oder sogar Gebäude verteilt geschrieben werden muss. Eine der regelmäßig größten Prüfungen führt dabei das Englische Seminar mit über 300 Teilnehmenden durch. Studieninteressierte können sich so ohne kostenpflichtige standardisierte Sprachtests auf einen Studienplatz bewerben.

Weitere größere Klausuren mit über 100 Teilnehmenden werden regelmäßig von der Fakultät für Umwelt und Natürliche Ressourcen abgehalten. Die meisten Prüfungen werden dabei als “E-Klausur plus” durchgeführt, da das Arbeiten mit der Statistik-Software RStudio Teil der Prüfungsleistung ist.

Im Jahr 2019 stieß die Technische Fakultät mit der “Einführung in die Informatik” zur “ü100”-Gruppe hinzu. Für diese Prüfungen wurde ein neuer Fragentyp eingesetzt, der es möglich macht, auch ohne die Freigabe von Drittsoftware Programmiercode zu schreiben und auszuführen (z. B. Python). Die bisher größte Prüfung erfolgte im März 2020 mit über 360 angemeldeten Teilnehmenden. Die Klausur wurde in zwei Kohorten über mehrere Gebäude verteilt geplant und erfolgreich durchgeführt. [He20]

### **E-Prüfungen mit besonderen Elementen**

Die Sportwissenschaft gehörte zu den ersten Fächern, die elektronische Prüfungen an der Uni Freiburg umsetzte. Die Attraktivität der elektronischen Umsetzung der Gerätturnen-Klausur bestand von Anfang an im Zeigen von Videos, bei denen die Studierenden Bewegungsabläufe analysieren müssen.

Auch die Medienkulturwissenschaften profitieren mittlerweile von elektronischen Prüfungen und schreiben regelmäßig Klausuren mit bwLehrpool und ILIAS. Hier haben vor allem Studierende den Vorteil, Gelerntes bei der Analyse von Filmausschnitten direkt praktisch umsetzen zu können.

E-Prüfungen mit besonderen Elementen erfordern von den Lehrenden einige Fähigkeiten und spezielle Vorbereitungen, doch durch ein gemeinsames Herangehen mit dem Klausuren-Team wurden diese nach und nach aufgebaut und optimiert. Fragen wurden verbessert und sowohl für Studierende als auch für Lehrende attraktiver gestaltet. [SR20]

## **3.2 Prüfung unter Nutzung originaler Softwareumgebungen**

Auf virtuellen Maschinen basierende E-Prüfungen bieten neue Anwendungsmöglichkeiten, da die gleiche Umgebung, die von Studierenden zum Lernen verwendet wurde, unabhängig vom Betriebssystem in gewohnter Arbeitsweise auch zur Prüfung eingesetzt werden kann. Hierbei lässt sich die bwLehrpool-Prüfungsumgebung je nach gewünschten Freiheitsgraden der Klausur spezifisch konfigurieren: Ist etwa

Netzzugriff zur Recherche gestattet? Oder darf eventuell nur auf bestimmte Server zugegriffen werden? Soll die Abgabe der Ergebnisse auf einen bestimmten Server hochgeladen werden, oder ist dafür ein passend berechtigtes Netzlaufwerk vorgesehen? Sind USB-Sticks - etwa zur Ausgabe der Aufgabenstellung und Abgabe der Ergebnisse - erlaubt, oder sollen diese ignoriert werden? Virtuell basierte E-Prüfungen befinden sich aufgrund der genannten Vorteile im Aufwind. Als E-Prüfung kann im Prinzip jede Softwareumgebung direkt verwendet werden, da der bwLehrpool-Prüfungsmodus die Sicherheit der Prüfung ohne Zugriffsmöglichkeit der virtuellen Maschine nach außen gewährleistet.

#### **4 Fazit und Ausblick**

Seit 2016 wurden an der Universität Freiburg mit dem aktuell betriebenen System bei steigender Nachfrage über 60 Prüfungen mit insgesamt über 3800 Prüflingen geschrieben. Viele schon an anderer Stelle beschriebene Vorteile von E-Assessments bestätigten sich auf breiter Linie. Zwar erfordert eine Umstellung auf E-Klausuren bei guter Gestaltung viel Zeit seitens der Lehrenden, allerdings steht diesen langen Vorbereitungsarbeiten eine deutliche Zeitersparnis bei der Korrektur gegenüber. Viele Klausuren sind eine Mischung aus automatisch auswertbaren und "normal" zu korrigierenden Fragen. Mit der digitalen Abgabe im LMS sind die Zeiten vorbei, in denen das Entziffern von Handschriften eine große Herausforderung bei der Korrektur darstellte. Die begutachtende Person sieht immer dieselbe Schriftart, was wiederum den Spielraum für subjektiv wahrgenommenes Können der zu Beurteilenden auf ein minimales Maß beschränkt. In vielen Fächern konnte dadurch die Zeit bis zur Rückmeldung der Ergebnisse an die Studierenden deutlich reduziert werden.

Große Vorteile haben E-Klausuren für Fächer, die mit Video, Audio oder hochauflösenden Bildern arbeiten, was sich bisher nur schwer in Klausuren umsetzen ließ. Die bisher eher praxisfernen Papier-Programmierklausuren können durch die Weiterentwicklung entsprechender Fragetypen in realistische Prüfungen in praxisnaher Arbeitsumgebung transformiert werden.

E-Prüfungen haben sich dank engagierter Zusammenarbeit aller Beteiligten bewährt. Die bwLehrpool-Prüfungsumgebung mit vorinstalliertem Safe Exam Browser bieten eine hohe Rechtssicherheit bei der Durchführung von Prüfungen. Inzwischen stößt der "Baden-Württemberg"-Lehrpool auch in weiteren Bundesländern und darüber hinaus auf großes Interesse. So wird beispielsweise in Nordrhein-Westfalen versucht, das System in enger Abstimmung als "lehrpool.nrw" anzubieten. Durch die günstige Lage im Dreiländereck bestehen zudem erste Anknüpfungspunkte zur Universität Straßburg, an der Universität Basel wird bwLehrpool ebenfalls bereits evaluiert und schrittweise eingeführt.

## Literaturverzeichnis

- [VS09] Vogt, M.; Schneider, S: E-Klausuren an Hochschulen: Didaktik – Technik – Systeme – Recht – Praxis, Giessen, 2009.
- [EL20] ELAN e.V. [https://ep.elan-ev.de/wiki/E-Klausur#Digitale\\_Ausf.C3.BCIIhilfe](https://ep.elan-ev.de/wiki/E-Klausur#Digitale_Ausf.C3.BCIIhilfe). [Zugriff am 06 April 2020].
- [Gi17] Gidion, G. et al.: Mediale Hochschul-Perspektiven 2020 in Baden-Wuerttemberg : empirische Untersuchung im Rahmen der Allianz "Forward IT", KIT Scientific Publishing, 2017.
- [Ri16] Ritter, S. et al.: bwLehrpool: Durchführung von elektronischen Prüfungen in virtualisierten Umgebungen, DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik, 2016.
- [DK19] Daniel M.; Köcher, N.: Praxisbuch eKlausur. Elektronisches Prüfen mit ILIAS, 2019.
- [He20] Heybrock, M.: Papierlos und individualisierbar. Viele Fachbereiche sehen den Vorteil von E-Klausuren, <https://www.pr.uni-freiburg.de/pm/online-magazin/lehren-und-lernen/papierlos-und-individualisierbar> [Zugriff am 22. Juni 2020]
- [SR20] Slotosch, S.; Rößler, C.: Elektronisch prüfen mit bwLehrpool: Praxisorientiert und multimedial, <https://www.lehre.uni-freiburg.de/notizblog-lehre/elektronisch-pruefen-mit-bwlehrpool-praxisorientiert-und-multimedial> [Zugriff am 22. Juni 2020]



## Eine Studie zur Qualitätsbeurteilung von automatisierten Testwerkzeugen zur Prüfung auf Barrierefreiheit

Saba Mateen<sup>1</sup>, Sarah Voß-Nakkour<sup>2</sup> und Linda Rustemeier<sup>3</sup>

**Abstract:** Damit digitale Bildungsangebote von allen Menschen mit und ohne Beeinträchtigungen gleichermaßen benutzbar sind, müssen diese barrierefrei vorliegen. Sollte dies nicht der Fall sein, können Nutzer\*innen, die z.B. ein Bildschirmlesegerät verwenden, nicht auf alle Inhalte zugreifen. Bei einer barrierefreien Aufbereitung helfen automatische Testwerkzeuge, welche Auskunft über vorliegende Barrieren geben. Diese Studie geht der Frage nach, wie die Qualität dieser Tools in Bezug auf Webseiten zu bewerten ist. Dazu werden sieben verschiedene Testwerkzeuge auf Webseiten mit unterschiedlicher Konformität der Web Content Accessibility Guidelines (WCAG) ausgeführt. Die resultierenden Ergebnisse werden in ihrer Vollständigkeit, Korrektheit und Spezifität untersucht und beschrieben. Ziel der Studie ist es, Entwickler\*innen und Web-Redakteur\*innen verlässliche Testwerkzeuge zu empfehlen, die sie in der Erstellung digitaler barrierefreier Inhalte zielführend verwenden können. Die Ergebnisse der Evaluation zeigen, welche Barrieren zuverlässig aufgedeckt werden und wieso eine vollständig automatische Überprüfung nach dem jetzigen Stand der Tools nicht ausreichend ist.

**Keywords:** Digitale Barrierefreiheit; Testwerkzeuge; Inklusion

### 1 Einleitung

Im September 2018 ist die EU-Richtlinie 2016/2102<sup>4</sup> in das deutsche Recht überführt worden. Laut dieser müssen bis September 2020 die Web-Angebote öffentlicher Stellen der europäischen Norm EN 301 549 entsprechen, um so digitale Barrierefreiheit gewährleisten zu können [He17]. Was unter digitaler Barrierefreiheit zu verstehen ist und welche weiteren rechtlichen Grundlagen gelten, wird im Folgenden dargelegt.

#### 1.1 Digitale Barrierefreiheit

Häufig wird mit dem Begriff „Barrierefreiheit“ unmittelbar eine Rollstuhllrampe, ein Treppenlift oder ein Blindenleitsystem assoziiert. Der tägliche Zugriff auf Webseiten und digitale Systeme ist für die Gesellschaft und im Bildungswesen selbstverständlich, sodass

---

<sup>1</sup> Goethe-Universität Frankfurt am Main, studiumdigitale, Varrentrappstr. 40-42, 60486 Frankfurt am Main, mateen@studiumdigitale.uni-frankfurt.de

<sup>2</sup> Goethe-Universität Frankfurt am Main, studiumdigitale, Varrentrappstr. 40-42, 60486 Frankfurt am Main, voss@studiumdigitale.uni-frankfurt.de

<sup>3</sup> Goethe Universität Frankfurt am Main, studiumdigitale, Varrentrappstr. 40-42, 60486 Frankfurt am Main, rustemeier@studiumdigitale.uni-frankfurt.de

<sup>4</sup> Richtlinie (EU) 2016/2102 des europäischen Parlaments und des Rates vom 26. Oktober 2016 über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen; Artikel 4

auch dort die Barrierefreiheit berücksichtigt werden muss. Tätigkeiten und Aktionen, die früher analog erledigt wurden, werden nun vermehrt digital bearbeitet. Das folgende Szenario verdeutlicht dies: Eine blinde Studentin möchte sich mithilfe eines Bildschirmlesegeräts in die Lernplattform einloggen, um eine Abgabe hochzuladen. Als sehende\*r Nutzer\*in erkennt man zwei Eingabefelder mit dem jeweils darüberstehenden Text „Benutzername“ und „Passwort“, welchen jedoch kein Label hinterlegt ist. Das Bildschirmlesegerät liest nur „Eingabefeld“ vor. Folglich ist es der Studentin nicht möglich, sich selbstständig einzuloggen, da ihr nicht vermittelt wird, welche Information wo gefordert ist. Das ist nur ein Beispiel für mögliche digitale Barrieren. Es gibt noch weitaus mehr wie dieses Paper zeigt. Allgemein kann man unter digitaler Barrierefreiheit, das Zugänglichmachen von digitalen Inhalten für alle Benutzer\*innengruppen, einschließlich derer mit körperlichen oder psychischen Beeinträchtigungen, verstehen (Unterkapitel 1.2). Barrierefreie Digitalangebote sollten kurz gesagt wahrnehmbar, bedienbar, verständlich und robust sein [W318b]. Ferner ist anzumerken, dass die Aufbereitung hinsichtlich digitaler Barrierefreiheit nicht nur Menschen mit Behinderungen und/oder Beeinträchtigungen zu Gute kommt, sondern dass alle Nutzer\*innen einen Vorteil durch diese Aufbereitung davon tragen. So helfen zum Beispiel das Hinzufügen von Untertiteln bei Videos nicht nur hörbeeinträchtigten oder gehörlosen Nutzer\*innen, sondern auch solchen, die sich ein Video ansehen möchten, aber in einer Situation sind, in welcher sie keinen Ton abspielen können. Auch das Ermöglichen von einer Textvergrößerung bis zu 200% hilft nicht nur sehbeeinträchtigten Nutzer\*innen, sondern auch denjenigen, die nach langem Arbeiten am Bildschirm die kleine Schrift nicht mehr gut erkennen oder auch älteren Menschen, deren Sehkraft nachlässt.

## 1.2 Gesetze, Verordnungen, Richtlinien, Normen und Empfehlungen

Auf der Gesetzesebene sowie in der EU-Richtlinie 2016/2102 wird die Beseitigung von Benachteiligungen angestrebt. Im Grundgesetz<sup>5</sup> (GG) und dem Allgemeinen Gleichbehandlungsgesetz<sup>6</sup> (AGG) wird die Benachteiligung aufgrund einer Behinderung untersagt. Das Behindertengleichstellungsgesetz<sup>7</sup> (BGG) und die EU-Richtlinie 2016/2102 klären zusätzlich, welche digitalen Inhalte öffentlicher Stellen bis wann barrierefrei aufbereitet werden müssen. Als Grundlage für die Erfüllung der genannten Rechtsprechungen müssen in Deutschland die Vorschriften aus der Barrierefreien-Informationstechnik-Verordnung<sup>8</sup> (BITV) beziehungsweise der Europäischen Norm 301 549<sup>9</sup> eingehalten werden. Sowohl die BITV wie auch die EN 301 549 beruhen auf den internationalen Web Content Accessibility Guidelines (WCAG) [He17]. Diese sind eine Sammlung von Empfehlungen, welche Barrierefreiheit von Webinhalten gewährleisten

<sup>5</sup> Grundgesetz (GG); Artikel 3 Absatz 3

<sup>6</sup> Allgemeines Gleichbehandlungsgesetz (AGG); § 1

<sup>7</sup> Behindertengleichstellungsgesetz (BGG); § 12a

<sup>8</sup> Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0); § 1 Ziele (1)

<sup>9</sup> EN 301 549 V1.1.2 (2015-04) - Accessibility requirements suitable for public procurement of ICT products and services in Europe

soll und durch die Accessibility Guidelines Working Group, einer Arbeitsgruppe des World Wide Web Consortiums (W3C) veröffentlicht worden ist. Unter den 179 Mitgliedern finden sich 27 eingeladene Accessibility-Expert\*innen sowie 152 Vertreter\*innen großer Firmen aus dem Bereich der Informationstechnik wie Google, Microsoft und Adobe, aber auch Vertreter\*innen aus Firmen, die sich auf digitale Barrierefreiheit spezialisiert haben [W317b]. Die Empfehlungen sind in vier Prinzipien unterteilt: Wahrnehmbarkeit, Verständlichkeit, Bedienbarkeit und Robustheit. Unter diesen werden 13 Richtlinien aufgeteilt und als Fundament der Barrierefreiheit für das Web beschrieben [He09]. Diese Richtlinien lassen sich in 78 Erfolgskriterien untergliedern [W318b]. 17 dieser Kriterien sind in der Version 2.1 im Juni 2018 hinzugekommen und verfolgen die Verbesserungen für Menschen mit Sehbeeinträchtigung, geistlicher Beeinträchtigung und Lernbehinderungen. Auch Barrieren auf mobilen Endgeräten werden mithilfe der Neufassung vermindert [W318a]. Im Februar dieses Jahres ist ein erster Entwurf der Version 2.2 erschienen, der ein neues Erfolgskriterium beinhaltet. Dieses beansprucht den sichtbaren Tastaturfokus zusätzlich als Mindestanforderung (Level A) [W320]. Unter Beachtung dieser Empfehlungen können Entwickler\*innen Webseiten für Menschen mit und ohne Beeinträchtigungen barrierefrei gestalten. Bei der Überprüfung können außerdem automatische Testwerkzeuge genutzt werden, welche Auskunft über vorliegende Barrieren geben. Die in diesem Paper beschriebene Studie geht der Frage nach, wie die Qualität ebensolcher Tools zu bewerten ist. Dazu werden im Folgenden sieben verschiedene Testwerkzeuge auf Webseiten mit unterschiedlicher Konformität der WCAG ausgeführt und ausgewertet.

## **2 Rahmenbedingungen der Studie**

Bevor die Testwerkzeuge untersucht werden, müssen zunächst Rahmenbedingungen definiert werden, sodass ein fundiertes Studienergebnis erzielt werden kann. Die Wahl der Tools und der Untersuchungsumgebung wird im Folgenden näher erläutert.

### **2.1 Auswahl der zu überprüfenden Webseiten**

Um die Webseiten bestmöglich zu testen und die Ergebnisse zu bewerten, sind die ausgesuchten Webseiten und Tools entscheidend. Dafür müssen Webseiten ausgewählt werden, auf deren Basis die Tools verschiedene Aspekte beleuchten können. Ein zu überprüfender Aspekt stellt die Ausgabe falsch-negativer Ergebnisse dar (Unterkapitel 3.1). An dieser Stelle bietet es sich an, eine Seite mit vielen Barrieren auszuwählen, sodass untersucht werden kann, ob das Testwerkzeug diese vollständig findet. Um zusätzlich die Ausgabe falsch-positiver Ergebnisse zu überprüfen (Unterkapitel 3.2), wird eine barrierearme Seite gewählt. Hierdurch wird geprüft, ob die Testwerkzeuge Barrieren ausgeben, die jedoch tatsächlich gar keine darstellen. Die Wahl der zu überprüfenden Webseiten fällt zum einen auf die Seite der Fachstelle für Ergänzende unabhängige

Teilhabeberatung<sup>10</sup>. Diese beruft sich in ihrer Barrierefreiheitserklärung auf die Durchführung eines BITV/WCAG-Selbstbewertungstest, welcher zeigt, dass die Webseite mit den entsprechenden Richtlinien konform ist. Ergänzend wird der Webauftritt des Projekts „Digital gestütztes Lehren und Lernen in Hessen“<sup>11</sup> betrachtet. Dieser verwendet das Wordpress Fluida-Theme und arbeitet derzeit in Kooperation mit dem Innovationsforum Barrierefreiheit<sup>12</sup> daraufhin dieses barrierefrei aufzubereiten. Fortführend wird eine Testwebseite unter Verwendung des gleichen Themes erstellt, jedoch mit der Intention bewusst viele, bekannte Barrieren einzubauen, die in Überprüfungen von nicht-barrierefreien Webseiten im Rahmen der Arbeit im „Innovationsforum Barrierefreiheit“ auffällig geworden sind. So kann geprüft werden, wie verlässlich die Testwerkzeuge eine breite Fächerung an häufig auftretenden Barrieren ausfindig machen und ob falsch-negative Ergebnisse zugelassen werden.

## 2.2 Auswahlkriterien der zu evaluierenden Tools

Die zu testenden Tools entsprechen den „Web accessibility evaluation tool[s]“, also Software Anwendungen, welche dazu entwickelt worden sind, um Webinhalte gemäß gewisser Forderungen an Barrierefreiheit, wie den WCAG, zu überprüfen [W317a]. Als Grundlage für die Auswahl an Tools wird zunächst die „Web Accessibility Evaluation Tools List“ [W316] des World Wide Web Consortiums (W3C) herangezogen. Die Intention hierbei ist es, solche Tools zu evaluieren, die durch eine zuverlässige Quelle aufgelistet werden. Die W3C sind ein Zusammenschluss aus Expert\*innen im Gebiet der Barrierefreiheit und gleichzeitig die Herausgeber der WCAG. Auch die Bundesfachstelle Barrierefreiheit<sup>13</sup> sowie BIK für Alle<sup>14</sup> verweisen auf die eben genannte Liste der W3C. Aus diesem Grund könnte man davon ausgehen, dass die Tools, die auf der genannten Liste aufgezählt werden, regelkonform funktionieren. Doch wie auf der Seite selbst angemerkt wird, handelt es sich um eine reine Auflistung existierender Tools, die keiner vorherigen Evaluation zugrunde liegen. Aus diesem Grund wird in diesem Beitrag evaluiert, wie die Qualität der Ergebnisse einiger Tools aus der Liste zu bewerten ist, um für Entwickler\*innen und Web-Redakteur\*innen Aussagen über die Verlässlichkeit der Testwerkzeuge bei der Erstellung barrierefreier digitaler Inhalte treffen zu können. Die 140 Tools werden auf jene eingegrenzt, welche auf die WCAG 2.1-Konformität prüfen, da diese, wie im Kapitel 1.2 behandelt, meist als Grundlage für gesetzliche Vorschriften gewählt werden. Nach dieser Eingrenzung bleiben 32 Tools zur Auswahl. Aus diesen fallen Tools raus, welche nicht zur Evaluation von Webseiten dienen. Testwerkzeuge wie Toolbars, die lediglich manuelle Prüfschritte unterstützen oder automatische Schnelltests auf häufig auftretende Barrieren durchführen, werden auch ausgeschlossen. Nach Durchführen des Ausschlussverfahrens werden nun die Tools „Access Continuum

<sup>10</sup> <https://www.teilhabeberatung.de>

<sup>11</sup> <https://www.digll-hessen.de>

<sup>12</sup> <https://www.digll-hessen.de/innovationsforen/innovationsforum-barrierefreiheit>

<sup>13</sup> <https://www.bundesfachstelle-barrierefreiheit.de/DE/Praxishilfen/Informationstechnik/Testen/testen.html>  
(Stand: 15.04.2020)

<sup>14</sup> <https://bik-fuer-alle.de/automatisierte-tests.html> (Stand: 15.04.2020)

Community Edition“ (ACCE)<sup>15</sup>, „ANDI“<sup>16</sup>, „Mauve“<sup>17</sup>, „TAW“<sup>18</sup> sowie „Wave“<sup>19</sup> gewählt. Ergänzend werden die Ergebnisse der Umfrage „Digitale Barrierefreiheit an hessischen Hochschulen“<sup>20</sup> betrachtet. Diese Umfrage ist im Rahmen des digLL-Projekts<sup>20</sup> zur Untersuchung des Bedarfes digitaler Barrierefreiheit an elf hessischen Hochschulen durchgeführt worden. Unter anderem ist abgefragt worden, welche Tools die Befragten für die barrierefreie Aufbereitung von Webseiten nutzen. Nach dem Heranziehen dieser Umfrageergebnisse und dem Anwenden der obigen Auswahlkriterien werden zusätzlich noch die Tools „axe“<sup>21</sup> sowie „aChecker“<sup>22</sup> untersucht.

### 2.3 Eingrenzung des Untersuchungsspektrums

Um das Überprüfen der Tools übersichtlich und nachvollziehbar zu gestalten, wird das Untersuchungsspektrum eingegrenzt. Es wird nicht die Kompatibilität mit Touchscreens oder von mobilen Geräten betrachtet, da für diese zusätzliche Anforderungen gelten und die Untersuchung dieser über den Rahmen des Beitrags hinausgeht. Da lediglich die Qualität der Tools selbst im Vordergrund steht und nicht die der Webseiten, werden diese nicht vollständig überprüft. Die Evaluierung durch die Tools beschränkt sich auf die Kontrolle der Startseite des jeweiligen Webauftritts. Weiterhin werden die Tools im Chrome-Browser ausgeführt, was jedoch keinen Einfluss auf die Ergebnisse hat, da die Tools selbst den vom Browser unabhängigen HTML-Code evaluieren.

## 3 Untersuchungsverfahren

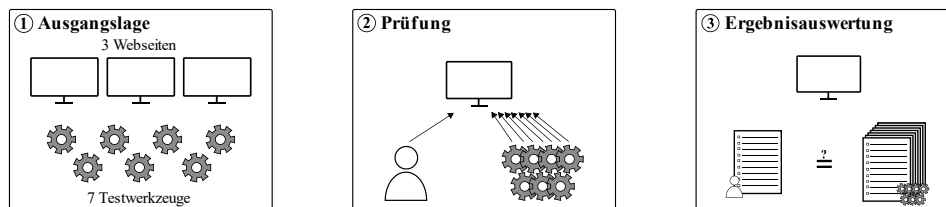


Abb. 1: Ablauf der Untersuchung

Wie die Ausgangslage zeigt (Abb. 1), werden sieben Testwerkzeuge auf drei Webseiten getestet. Im nächsten Schritt (Prüfung) wird jede Webseite manuell mit zuvor erlangter Expertise durch Kooperationen mit Einrichtungen wie dem Zentrum für blinde und

<sup>15</sup> <https://www.webaccessibility.com>

<sup>16</sup> <https://www.ssa.gov/accessibility/andi/help/install.html>

<sup>17</sup> <https://mauve.isti.cnr.it>

<sup>18</sup> <https://www.tawdis.net/?lang=en>

<sup>19</sup> <https://wave.webaim.org>

<sup>20</sup> <https://www.digll-hessen.de/innovationsforen/innovationsforum-barrierefreiheit>

<sup>21</sup> <https://www.deque.com/axe>

<sup>22</sup> <https://achecker.ca/checker/index.php>

sehbehinderte Studierende (BliZ)<sup>23</sup> im Rahmen der Mitarbeit im Projekt „Innovationsforum Barrierefreiheit“, in mehreren Stufen überprüft. Anschließend wird jedes Tool auf der Webseite ausgeführt. Hierbei wird darauf geachtet, dass die Tools mit einer möglichst geringen zeitlichen Differenz ausgeführt werden, um Abweichungen in den Testergebnissen aufgrund aktualisierter Webinhalte zu vermeiden. Nachdem die Überprüfung einer Seite durch alle Tools durchgeführt worden ist, werden im letzten Schritt (Ergebnisauswertung) die Ergebnisberichte auf Vollständigkeit (Unterkapitel 3.1), Korrektheit (Unterkapitel 3.2) und Spezifität (Unterkapitel 3.3) verglichen.

Im ersten Durchgang der manuellen Prüfung führt das Tool „Accessibility Insights for Web“<sup>24</sup> automatische Kontrollabfragen aus, wie die Suche nach Kontrastproblemen oder fehlenden alt-Attributen. Diese Ergebnisse werden nur zur Kontrolle der Vollständigkeit der rein manuellen Überprüfung herangezogen. Zum Überprüfen der Farbkontraste wird das Tool „Colour Contrast Analyser“<sup>25</sup> genutzt. Anschließend wird mit dem „Accessibility Insights for Web“ Tool weitergearbeitet, da das Tool die Nutzer\*innen auch bei manuellen Prüfschritten unterstützt. Auf Grundlage der WCAG 2.1 werden nacheinander manuelle Prüfschritte mit Erklärungen zur Durchführung aufgezählt. Hierzu bietet das Tool ergänzende Funktionen wie die Darstellung der Seite in Grautönen, Hervorhebung der Überschriften mit Anzeige der Überschriftenebenen, das Anzeigen von „Landmark Roles“ sowie das visuelle Hervorheben der Tab-Reihenfolge. Anschließend wird die BITV/WCAG-Selbstbewertung<sup>26</sup> durchgeführt. Unterstützend werden hierzu die Bookmarklets „Text Spacing“ wie auch „Inhalte gegliedert“ verwendet, welche vom BITV/WCAG Test empfohlen werden [BI19]. Als abschließenden Prüfschritt wird durch jede Seite mit den kostenfreien Screenreader NVDA<sup>27</sup> navigiert, um Barrieren aufzudecken, die durch die bisherigen Prüfschritte übersehen wurden.

### 3.1 Vollständigkeit

Um Tools auf Vollständigkeit zu testen, werden die automatisch gefundenen Barrieren jedes Tools jeweils mit den tatsächlich vorhandenen Barrieren aus der manuellen Überprüfung verglichen. Die Schnittmenge beider Ergebnismengen gibt Auskunft über die Qualität im Hinblick auf Vollständigkeit. Je größer die Schnittmenge ist, desto besser ist die Vollständigkeit des Testwerkzeugs. Mit anderen Worten: Je weniger falsch-negative Ergebnisse (siehe Abb. 2) vorliegen, desto vollständiger ist das Ergebnis. Falsch-negative Ergebnisse sind in diesem Fall vorhandene Barrieren, die von dem Tool nicht als solche erkannt werden. Solche Resultate haben schwere Konsequenzen, da sie den Nutzer\*innen des Tools nicht auf das Problem hinweisen, wodurch eine Barriere weiterhin unbemerkt als Teil der Seite fortbesteht. Der einzige Weg um falsch-negative Ergebnisse zu vermeiden, ist nach Brajnik das Einsetzen von weiteren Überprüfungsverfahren [Br04].

---

<sup>23</sup> <https://www.thm.de/bliz>

<sup>24</sup> <https://accessibilityinsights.io>

<sup>25</sup> <https://developer.paciellogroup.com/resources/contrastanalyser>

<sup>26</sup> <https://testen.bitv-test.de/selbstbewertung/index.php>

<sup>27</sup> <http://nvda.bhvd.de>

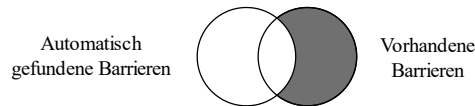


Abb. 2: Der graue Bereich entspricht den falsch-negativen Ergebnissen

### 3.2 Korrektheit

Die Korrektheit der Tools wird durch die Mengendifferenz der automatisch und manuell gefundenen Ergebnisse untersucht. Bei Barrieren, die nur von den automatischen Testwerkzeugen entdeckt worden sind, die jedoch keine tatsächliche Barriere darstellen, liegt ein falsch-positives Ergebnis (siehe Abb. 3) vor und die Korrektheit ist nicht erfüllt. Nach Brajnik ist die Konsequenz hier, dass das Tool nicht weiterverwendet wird, da es unnötige Störsignale aufwirft [Br04].

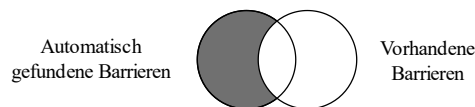


Abb. 3: Der graue Bereich entspricht den falsch-positiven Ergebnissen

### 3.3 Spezifität

Bei der Untersuchung nach der Spezifität der Tools wird die Genauigkeit der Toolergebnisse betrachtet. Wird beispielsweise ein „Spacer-Bild“<sup>28</sup> mit fehlendem Alternativtext gefunden, so ist die Fehlermeldung „Alternativtext fehlt“ wenig spezifisch. Besser sei nach Brajnik „Spacer-Bildern wird ein leerer alt-Tag zugewiesen“. So wird zwischen ähnlichen Barrieren unterschieden [Br04]. Zusätzlich wird die Ausführlichkeit der Erklärung einer Barriere betrachtet. Denn je ausführlicher ein Problem erklärt ist, desto spezifischer kann man verstehen, um welche Art von Barriere es sich handelt.

## 4 Ergebnisse

Da die Untersuchung der einzelnen Webseiten in dieser Studie kein Hauptaugenmerk darstellt, werden die vorhandenen Barrieren nicht seitenweise betrachtet, sondern zusammengefasst in einer Tabelle. Die aufgefundenen Barrieren sind entsprechend dem übergeordneten Erfolgskriterium gegliedert. Hat ein Tool nicht alle Vorkommen einer Barriere gefunden, so ist das Tool in der entsprechenden Spalte nicht angekreuzt worden. Hat ein Tool beispielsweise nur zwei der tatsächlich fünf fehlenden Alternativtexte gefunden (Tab. 1, Kriterium 1.1.1), so ist es in der folgenden Tabelle nicht angekreuzt.

<sup>28</sup> Bild zum Erzeugen von Leerflächen

WCAG-Erfolgskriterium	Identifizierte Barrieren	aChecker	ACCE	ANDI	axe	Mauve	TAW	Wave
1.1.1	Fehlender Alternativtext	x	x	x	x		x	x
	Unzureichender Alternativtext							
1.3.1	CSS-Code zum Ein- und Ausblenden von Inhalt	x			x	x	x	x
	Falsche Auszeichnung von Textstruktur				x			x
	Fehlende Auszeichnung von Textstruktur							
	Fehlende Überschrift der ersten Ebene	x						
	Übersprungene Überschriftenebene(n)			x				
1.4.1	Informationsvermittlung durch Farbe							
1.4.3	Kontrastproblem Text			x	x			x
1.4.5	Verwendung einer Textgrafik							
1.4.11	Kontrastproblem Grafiken							
1.4.13	Hover-Inhalt kann nicht verworfen werden							
2.1.1	Inhalt nicht ohne Benutzung der Maus zu erreichen			x				
2.4.2	Titel der Webseite fehlt	x	x	x	x	x		x
2.4.4	Link nicht ohne Kontext verständlich							
3.1.2	Fehlende Sprachauszeichnung eines Wortes/Absatzes							
3.3.2	Eingabefeld ohne Kennzeichnung (Label)	x	x		x	x	x	x
Anzahl gefundener Barrieren von 17		5	3	5	6	4	3	6

Tab. 1: Identifizierte Barrieren der getesteten Webseiten und Suchergebnisse der Tools



## 5 Auswertung der Ergebnisse

In diesem Kapitel werden die Ergebnisse hinsichtlich der in Kapitel 3 vorgestellten Methoden nähergehend untersucht und bewertet, um so die Qualität der Testwerkzeuge beurteilen zu können.

### 5.1 Vollständigkeit

Im Durchschnitt haben die ausgewählten Tools knapp ein Viertel der vorhandenen Barrieren erfasst. Vor Allem sind technische Barrieren gefunden worden, die durch das Fehlen eines HTML-Attributs oder -Tags begründet werden können. So haben fünf von sieben Tools das Fehlen des <title>-Tags, <h1>-Tags und <label>-Tags sowie das Fehlen des ‚alt‘-Attributs gefunden. Besonders gut schneiden hier Wave und axe ab, welche etwa ein Drittel der vorliegenden Barrieren gefunden haben. Gegenteilig dazu ist das Ergebnis von ACCE und TAW, welche nur gut ein Sechstel der Barrieren ermittelt haben. Doch selbst wenn man die Ergebnisse der Tools zusammengefasst betrachtet, wird nur etwa die Hälfte aller Barrieren gefunden. Unter den nicht erkannten Barrieren handelt es sich vor Allem um solche, die eine Auswertung des Inhalts benötigen. So ist zum Beispiel von keinem Tool erkannt worden, dass Informationen mit Hilfe von Farben vermittelt werden oder ein Linktext ohne den Kontext nicht verständlich ist. Es gibt auch Barrieren, die nur von manchen Tools erkannt worden sind. Gründe dafür sind zum einen die unterschiedlich fortgeschrittenen Entwicklungsstände oder das Verständnis bzw. die Interpretation der WCAG. In dessen Empfehlungen steht zum Beispiel nicht, dass kursive Schrift zwingend mit dem <em>-Tag anstelle des <i>-Tags ausgezeichnet werden muss. Doch zum Erfüllen der Richtlinie 1.3.1 ‚Infos and Relationships‘ wird die Technik H49<sup>29</sup> benötigt, welche besagt, dass nur die Verwendung des <em>-Tags versichert, dass ein Bildschirmlesegerät programmtechnisch erkennen kann, ob Inhalte in kursiver Form dargestellt sind. Diese Barriere ist nur vom aChecker ausgegeben worden, obwohl diese durch tiefergehendes Wissen und Parsen<sup>30</sup> des HTML-Codes einfach ermittelt werden kann.

### 5.2 Korrektheit

Eine unzureichende Leistung unter dem Aspekt der Korrektheit zeigt das Tool Mauve auf. Dieses evaluiert Webseiten, indem es ausgewählte Techniken<sup>31</sup> überprüft. So gilt zum Beispiel für die korrekte Einbettung eines Links, dass dieser unter anderem entweder einen aussagekräftigen Linktext besitzt oder eine Auszeichnung mithilfe eines ARIA-Labels erfolgen muss. Ersteres ist durch alle Links der digLL-Seite erfüllt, in Mauve ist jedoch trotzdem für jeden Link der Fehler ausgegeben worden, dass kein ARIA-Label vorliegt.

<sup>29</sup> H49: Using semantic markup to mark emphasized or special text (<https://www.w3.org/TR/WCAG20-TECHS/H49.html> Stand 15.04.2020)

<sup>30</sup> maschinenlesbare Daten analysieren, segmentieren und codieren (<https://www.duden.de/rechtschreibung/parsen> Stand: 01.04.2020)

<sup>31</sup> <https://www.w3.org/WAI/WCAG21/Techniques>

So kommt es dazu, dass ein Seitenelement eine Fehlermeldung fälschlicherweise hervorruft, was zu falsch-positiven Ergebnissen führt. Zusätzlich sind einige dieser Ergebnisse nicht nachvollziehbar. Es wird ausgegeben, dass die Sprachauszeichnung des HTML-Dokuments fehlt, was jedoch nicht der Fall ist. Auch werden mehr Kontrastprobleme aufgezählt als tatsächlich vorhanden sind. Die Ergebnisse, die möglicherweise eine Barriere darstellen, jedoch nicht zwingend müssen, werden durch die Tools aChecker, axe und Wave in die Kategorie der potenziellen Probleme zusammengefasst. Auf diese Art können die Nutzer\*innen die richtig-positiven Ergebnisse klar von den falsch-positiven Ergebnissen trennen.

### **5.3 Spezifität**

ANDI, axe und Wave haben hier am besten abgeschnitten. Die Tools markieren Barrieren visuell auf der Webseite, sodass ersichtlich ist, um welches Seitenelement es sich handelt. Zusätzlich hat man die Möglichkeit, sich die barrierenbehaftete Stelle im Quellcode anzeigen sowie sich eine Erklärung ausgeben zu lassen, welche verdeutlicht, wieso diese Barriere behoben werden sollte und wie man dabei vorgehen kann. Die Tools aChecker, Access Continuum Community Edition (ACCE) sowie TAW geben zu jeder ermittelten Barriere an, in welcher Codezeile diese zu finden ist, sodass Entwickler\*innen überblicken können, an welcher Stelle man ansetzen muss, um eine Barriere zu beheben. Nutzer\*innen des Tools, die kein (Vor-)wissen bezüglich des Programmierens von Webseiten haben, können diese Darstellung der Ergebnisse eventuell nicht als spezifisch genug beurteilen. Die Ergebnisausgabe des Tools Mauve gibt den Nutzer\*innen lediglich eine Auflistung der gefundenen Barrieren ohne Verweis auf die Webseite. So ist unklar, wo eine Barriere vorliegt und muss selbst teilweise zeitintensiv gesucht werden.

### **5.4 Automatisiertes vs. Manuelles Prüfen**

Automatische Testwerkzeuge geben Nutzer\*innen innerhalb weniger Sekunden eine Liste potenzieller Barrieren aus, wohingegen sich eine sorgfältige manuelle Prüfung über Stunden hinausziehen kann. Doch trotz des deutlich höheren Aufwandes sollte man nicht auf das manuelle Überprüfen verzichten. Wie im Unterkapitel 5.1 bereits dargelegt wurde, gibt es Barrieren, die von den Tools nicht erkannt werden. Hierzu gehören unter anderem unzureichende Alternativtexte, fehlende Auszeichnung von Textstrukturen, Informationsvermittlung durch Farbe, Links, die nicht ohne Kontext verständlich sind sowie die fehlende Sprachauszeichnung eines Wortes bzw. Absatzes. Eine Erklärung für diesen Sachverhalt gibt die Initiative „Einfach für Alle“ [TC]: „Für eine Software ist nicht feststellbar, wenn bestimmte Elemente einer Sprache missbräuchlich eingesetzt werden. Die Barrierefreiheit auf der Ebene der Wahrnehmung wie eingängige Navigation und verständliche Texte ist nur durch Tests mit echten Benutzern überprüfbar.“ Auch „BIK für alle“ [BI] bietet für diese Aussage ein bestärkendes Beispiel: „In der BITV [wie auch der WCAG] ist festgelegt, dass Bilder [...] mit Textalternativen zu versehen sind. Automatische Tools erkennen, ob Alternativen in Textform bereitgestellt werden. Sie

erkennen aber nicht, ob diese Texte tatsächlich eine sinnvolle Beschreibung [...] liefern. Das können nur Menschen beurteilen.“ Bei den verbleibenden nicht-aufgedeckten Barrieren handelt es sich um Kontrastprobleme von Grafiken und die Verwendung von Textgrafiken. Dieses Ergebnis lässt sich durch fehlende Auszeichnung im Quellcode erklären. Grafiken werden mit dem `<img>`-Tag eingefügt, welcher keine Auskunft über die vorhandenen Farben der Grafik gibt und auch nicht, ob eine Textgrafik vorliegt.<sup>32</sup> An dieser Stelle wird eine genauere Auszeichnungsmöglichkeit in Programmiersprachen benötigt oder eine visuelle Erkennung von Farbe und Text. Bevor diese von den Tools umgesetzt wird, müssen diese Eigenschaften von manuellen Prüfer\*innen erfüllt werden.

## 6 Fazit und Ausblick

Nach systematischer Prüfung der drei Webseiten mit den sieben Tools ist festzuhalten, dass axe und Wave quantitativ und qualitativ die besten Ergebnisse erreichen. An nächster Stelle folgt das Tool ANDI, welches ebenfalls verlässliche Resultate ausgegeben hat. Zwar hat aChecker die gleiche Menge an Barrieren gefunden, doch die aussagekräftige Visualisierung der Ergebnisse durch ANDI ist hinsichtlich der Nachvollziehbarkeit empfehlenswerter. Die Tools Mauve, TAW und ACCE sind basierend auf den Ergebnissen dieser Untersuchung weniger empfehlenswert, da diese keine visuelle Möglichkeit zum Nachvollziehen der Ergebnisse bieten und eine geringere Anzahl an richtig-positiven Ergebnissen ausgeben. Doch selbst die Tools mit den besten Ergebnissen haben nicht alle Barrieren identifiziert. Im Rahmen dieser Studie ist nur eine Auswahl an Tools, nach vorheriger Festlegung der Kriterien, getestet worden. Mit Sicherheit gibt es weitere Tools, die ähnliche oder bessere Ergebnisse liefern. Nichtsdestotrotz lässt sich festhalten, dass auch diese nicht alle Barrieren auffinden, da wie in Kapitel 5.4 gezeigt, das manuelle Prüfen nicht ersetzt werden kann. So ist es zum aktuellen Zeitpunkt nicht möglich, sich bei der Untersuchung von Webseiten vollkommen auf Testwerkzeuge zu verlassen. Barrieren, die zum Beispiel eine Auswertung des Inhalts benötigen, werden im Laufe der Untersuchung von keinem Tool erkannt. Hierunter fallen Probleme, wie die Informationsvermittlung durch Farbe sowie ein Linktext, der ohne Kontext nicht verständlich ist. Trotzdem kann das Verwenden solcher Tools die Nutzer\*innen zeitsparend und effizient unterstützen, vorausgesetzt man arbeitet mit einem verlässlichen Tool, welches keine aufwändige Filterung der Ergebnisse nach falsch-positiven Ausgaben benötigt und auskunftsreiche Fehlermeldungen bietet. In weiterführenden Arbeiten wird der Fokus auf Lernplattformen wie „Moodle“<sup>33</sup> und deren barrierearme Gestaltung gelegt, um so in Zukunft einen inklusiven Bildungsalltag zu ermöglichen. Auch weitere Aspekte müssen überprüft werden. Hierunter fällt beispielsweise die Usability der Testwerkzeuge. Es muss betrachtet werden, ob die Nutzung der Tools für alle Benutzer\*innen verständlich gestaltet ist und barrierefrei zu bedienen sind. Auch der zeitlich optimale Einsatz solcher

---

<sup>32</sup> [https://www.w3schools.com/tags/tag\\_img.asp](https://www.w3schools.com/tags/tag_img.asp) (Stand: 15.04.2020)

<sup>33</sup> <https://moodle.org/?lang=de>

Tools im Prozess der Webseitenentwicklung und deren potenzielle technische Verbesserung zum Beispiel unter Einbeziehung von KI für das Kontrollieren inhaltlich unzureichender Alternativtexte sind Themen weiterer Arbeiten. Zudem gibt es neben den überprüften Arten von Testwerkzeugen auch solche, die sich primär auf eine Evaluierung spezialisieren. Beispiele wären HeadingsMap<sup>34</sup>, welches sich auf die Untersuchung von Überschriftenebenen fokussiert oder A11y Color Contrast Accessibility Validator<sup>35</sup>, welches eine Webseite auf Kontrastprobleme überprüft. Es gilt zu evaluieren, ob Tools, die sich speziell auf eine Untersuchung begrenzen, gegebenenfalls verlässlichere Ergebnisse liefern als solche, die versprechen verschiedene Barrieren aufzufinden.

## Literaturverzeichnis

- [BI] BIK.: Automatisierte Tests. url: <https://bik-fuer-alle.de/automatisierte-tests.html>. Stand: 27.03.2020.
- [BI19] BIK.: Werkzeugliste des BITV/WCAG Tests. 2019. url: [https://www.bitvtest.de/bitv\\_test/das\\_testverfahren\\_im\\_detail/werkzeugliste.html](https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/werkzeugliste.html). Stand: 19.03.2020.
- [Br04] Brajnik, G.: "Comparing Accessibility evaluation tools: a method for tool effectiveness". Springer-Verlag. In: (2004).
- [Ca] Caspers, T.: Werkzeuge zur Überprüfung ihrer Webseite. url: <https://www.einfach-fuer-alle.de/artikel/test-werkzeuge>. Stand: 18.03.2020.
- [He09] Henry, S.: How WCAG 2.0 Differs from WCAG 1.0. 2009. url: <https://www.w3.org/WAI/WCAG20/from10/diff.php>. Stand: 14.03.2020.
- [He17] Hellbusch, J.: EN 301 549 als Mindestanforderung. 2017. url: <https://www.hellbusch.de/en-301-549-als-mindestanforderung/>. Stand: 15.03.2020.
- [W316] W3C.: Web Accessibility Evaluation Tools List. 2016. url: <https://www.w3.org/WAI/ER/tools/?q=wcag-21-w3c-web-content-accessibility-guidelines-21>. Stand: 18.03.2020.
- [W317a] W3C.: Developers' Guide to Features of Web Accessibility Evaluation Tools. 2017. url: <https://www.w3.org/TR/WAET/>. Stand: 14.03.2020.
- [W317b] W3C.: Participants in the Accessibility Guidelines Working Group. 2017. url: <https://www.w3.org/2000/09/dbwg/details?group=35422&public=1>. Stand: 14.03.2020.
- [W318a] W3C.: Comparison with WCAG 2.0. 2018. url: <https://www.w3.org/TR/WCAG21/#comparison-with-wcag-2-0>. Stand: 14.03.2020.
- [W318b] W3C.: Web Content Accessibility Guidelines (WCAG) 2.1. 2018. url: <https://www.w3.org/TR/WCAG21/>. Stand: 27.03.2020.
- [W320] W3C.: Web Content Accessibility Guidelines (WCAG) 2.2. 2020. url: <https://www.w3.org/TR/WCAG22/>. Stand: 27.03.2020.

<sup>34</sup> <https://addons.mozilla.org/es/firefox/addon/headingsmap>

<sup>35</sup> <https://color.a11y.com/?wc3>

# Nutzerzentrierter Ansatz zur Behebung von digitalen Barrieren

## Ein Entwicklungsframework für barrierefreie User Interfaces

Tornike Giorgashvili<sup>1</sup> und Sarah Voß-Nakkour<sup>2</sup>

**Abstract:** Die Prüfung von Webseiten auf Barrierefreiheit erfolgt nach wie vor noch unzureichend. Das Fehlen von geeigneten Standardtestverfahren ist einer der Gründe, warum Barrierefreiheit im Web so schwierig zu erreichen ist. Dieser Beitrag beschreibt, welche technischen Lösungsansätze sich mittels eines nutzerorientierten Ansatzes erstellen lassen, die automatisch möglichst viele digitalen Barrieren in einer Web-Anwendung abbauen. Es wird ein Entwicklungsframework für barrierefreie User Interfaces vorgestellt, das iterativ erarbeitet wurde. Neben den gewonnenen Erkenntnissen und entwickelten Ideen für technische Lösungen wird auch auf die Methode selbst eingegangen. Das entwickelte Framework AUI wurde entworfen, um für Nutzer\*innen einen „barriereärmeren“ Zugang zu digitalen Inhalten u.a. zu eLearning-Angebote zu ermöglichen. Mit einem funktionsübergreifenden Design liefert dieses Framework überschaubare Kernfunktionen, die u.a. von Menschen mit Behinderungen benötigt werden. Das AUI enthält spezifische Merkmale, die auf individuelle Bedürfnisse des Einzelnen abgestimmt werden können, um einer Person mit sensorischen, kognitiven oder körperlichen Behinderungen den Zugang zu einer Web-Anwendung zu ermöglichen.

**Keywords:** Barrierefreiheit, grafische Benutzeroberfläche, benutzerzentriertes Design, Framework, nutzerzentrierter Ansatz, Design Thinking

## 1 Einleitung

Digitales Lernen ist mittlerweile für die Gesellschaft eine wichtige Bildungsgrundlage. In einer zunehmend digitalisierten und global vernetzten Welt ist es deshalb wichtig, auch Menschen einen Zugang zu digitalen Lernangeboten zu ermöglichen, die aufgrund einer bestehenden (nicht selten) altersbedingten Seh- oder Hörbeeinträchtigungen, keinen oder nur einen eingeschränkten Zugang zu digitalen Lehrmaterialien haben [Ve18]. Für alle Lernenden mit oder ohne Beeinträchtigungen ist es aber notwendig, digitale Inhalte zu nutzen, um gesellschaftlich und wissenschaftlich teilnehmen zu können. Dies bedeutet gleichermaßen die Erstellung digitaler Lerninhalte, um eine digitale Barrierefreiheit herzustellen und damit eine uneingeschränkte Nutzung der digitalen Materialien. Jedoch stößt man bisher als Lernender bei digitalen Angeboten auf diverse Barrieren, die den

---

<sup>1</sup> Goethe-Universität Frankfurt am Main, studiumdigitale, Robert-Mayer-Str. 10, 60325 Frankfurt am Main, giorgashvili@studiumdigitale.uni-frankfurt.de

<sup>2</sup> Goethe-Universität Frankfurt am Main, studiumdigitale, Varrentrappstr. 40-42, 60486 Frankfurt am Main, voss@studiumdigitale.uni-frankfurt.de

Zugang zu Informationen oder E-Learning Angebote erschweren oder gar unmöglich machen, wenn man als Lerner\*in durch Beeinträchtigungen beschränkt ist beim Zugang von digitalem Lernmaterial.

Ziel muss es daher sein, digitale Barrieren zu vermindern oder ganz abzubauen, um, die Interaktion eingeschränkter Menschen mit einer Benutzungsoberfläche (User Interface (UI)) zu ermöglichen. Dem UI einer Webseite kommt dabei eine Schlüsselfunktion zu, da dieses der Nutzende die Interaktion mit den Funktionen einer Web-Anwendung ermöglicht. Digitale Barrieren in einem UI können in den meisten Fällen vermieden werden, wenn beim Design und während der Entwicklung ein paar wenige, für die Barrierefreiheit im Internet relevante Grundregeln Berücksichtigung finden. Gerade Entwicklern sind diese Grundregeln jedoch weitestgehend fremd. Auch die Notwendigkeit barrierefreier Webseiten wird nur von wenigen Entwickler\*innen verstanden [La13]. Digitale Barrieren bestehen nicht nur für seh- und hörgeschädigte Menschen. Auch nicht eingeschränkte Personen können durch technische Probleme daran gehindert werden, Internetangebote zu nutzen, etwa, wenn ihnen kein passender Browser, kein genügend schnelles Netzwerk oder kein Endgerät mit der benötigten Bildschirmauflösung zur Verfügung steht. Um den Anforderungen der Barrierefreiheit im Internet gerecht zu werden, hat das World Wide Web Consortium (W3C) verschiedene Richtlinien, die sog. Web Content Accessibility Guidelines (WCAG) 2.1. entwickelt, die den weltweiten gültigen Standard für barrierefreie Webdesign bilden [WC19]. Doch genügt es nicht nur, diese Standards einfach nur einzuhalten [HP11]: Die Umsetzung von Barrierefreiheit im Web ist ein fortlaufender Prozess [Me07]. Sie erfordert das Interesse und die Auseinandersetzung von Führungskräften und Webentwickler\*innen mit den besagten Richtlinien, woran es derzeit allerdings scheitert, weswegen noch immer viele Web-Angebote im Internet nicht barrierefrei sind [LO09]. Da eine nachträgliche Umrüstung von Web-Anwendungen auf Normen der Barrierefreiheit aufwändig und kostspielig ist, sollte vom Beginn des Entwicklungsprozesses an das Webdesign barrierefrei gestaltet werden. Es stehen zwar diverse Unterstützungstechnologien (UT) für Menschen mit Behinderung zur Verfügung, doch zeigen Studien, dass gerade die, die davon profitieren könnten, diese Technologien nicht nutzen [Ve18]. Das kann darauf zurückzuführen sein, dass der Einsatz von UT die Behinderung einer Person besser sichtbar macht. Darüber hinaus fühlen sich viele ältere Menschen nicht behindert, stellen aber fest, dass die Schriften kleiner sind, der Kontrast geringer ist oder der Ton weniger laut ist als früher.

Der folgende Beitrag hat deshalb zum Ziel, über die WCAG 2.1-Standards hinaus ein UI-Framework für Webentwickler\*innen und Webdesigner\*innen zu konzipieren, das digitale Barrieren automatisch abbaut, einen „barriereärmeren“ Zugang zu UIs ermöglicht und zudem universell in jede beliebige Web-Anwendung integrierbar ist. Das UI-Framework soll dabei Navigation und Interaktion mit einem UI für Menschen mit und ohne Behinderung gleichermaßen vereinfachen. Die in diesem Paper beschriebene Arbeit geht den Fragen nach, wie ein universell einsetzbares und zugängliches UI-Framework mittels nutzerzentrierten Ansatzes zu erarbeiten ist, wie dieses aussehen kann und welche Funktion es mit sich bringt.

## 2 Methodologie

Um bei der Konzeption des UI-frameworks den Anforderungen der Zielgruppe gerecht zu werden, wurde Design Thinking als nutzerzentriertes Vorgehen (User-Centered-Design (UCD)) gewählt. Dies hat ermöglicht, reale Nutzer\*innen mit unterschiedlichen Beeinträchtigungen in den wichtigsten Iterationen des Entwicklungsprozesses miteinzubeziehen. Der intensive Wissensaustausch mit den Teilnehmer\*innen hat dazu geführt, unterschiedlichen Prototypen zu erstellen und diese auf Barrierefreiheit zu testen. Anhand der erstellten Prototypen wurde ein Entwicklungsframework für „barriereärmere“ Benutzungsoberflächen erarbeitet, welches das Ergebnis dieser Arbeit darstellt.

### 2.1 Design Thinking

Design Thinking (DT) ist ein nutzerzentrierter Ansatz und verfolgt vor allem das Ziel, bei der Entwicklung der Lösungsansätze sich konsequent an den Bedürfnissen bzw. an den Anforderungen der Zielgruppe zu orientieren. Der DT-Prozess ist strukturiert, erfolgt iterativ und ist somit Kern der DT-Methode. Vor allem die häufigen Iterationen in dem Prozess dienen zur Verbesserung der Lösungen und ermöglichen damit einen Wissenszuwachs. [Sc17].

Durch Iteration des Vorgehens wird die jederzeit mögliche Rückkehr zur Vorgängerphase und damit die Anpassung bestimmter Parameter noch während des laufenden Prozesses ermöglicht. Der gewählte DT-Prozess fußt auf dem Ansatz Plattners und besteht insgesamt aus den folgenden sechs Phasen [Sc17]:

1. „Verstehen“: Problem und Problemkontext werden grundlegend erarbeitet.
2. „Beobachten“: Durch Beobachtung, Interviews etc. wird Kontakt zur Zielgruppe hergestellt.
3. „Standpunkt definieren“: Die in Phase 1 und 2 gesammelten Erkenntnisse werden evaluiert und gewichtet.
4. „Ideen finden“: In kurzer Zeit werden möglichst viele Ideen gesammelt.
5. „Prototypen“: Einfache Prototypen (Pappmodelle etc.) werden erstellt, um abstrakte Ideen möglichst früh zu visualisieren.
6. „Testen“: Die Prototypen werden von den späteren Nutzer\*innen getestet.

In dem DT-Prozess wird auf explizite Trennung zwischen Problemverständnis und Problemlösung geachtet [VE12], [Dr14]. Innerhalb des Prozesses sind Iterationen ausdrücklich erwünscht, denn durch schnelle Iterationen werden baldmöglichst Anregungen zu Lösungsvorschlägen gesammelt, um auf dessen Basis Verbesserungen der bisherigen Lösungen zu ermöglichen.

DT stellt die Nutzerorientierung und die Frage nach der Zielgruppe in den Vordergrund. Dies hilft insbesondere dabei, das Problem aus den Augen der Endbenutzer\*innen zu betrachten und Empathie aufzubauen, um somit ihre Bedürfnisse zu erkennen. Vor allem ist Empathie wichtig, um geeignete Lösungsstrategien für vorhandene Problemstellungen zu entwickeln [Sc17]. Es wird kontinuierlich sichergestellt, dass sowohl das Problem, als auch die Lösung aus der Perspektive der potenziellen Zielgruppe betrachtet werden. Dadurch wird sichergestellt, dass DT gute bedürfnisorientierte Ergebnisse liefert.

## 2.2 Teilnehmer\*innen

An allen Iterationen nahm eine Gruppe von sechs Testpersonen (T1 bis T6) teil. Diese Personen wurden durch persönliche Kontakte angeworben. Sie wurden zur Teilnahme ausgewählt, weil sie Grundwissen im Bereich der Informations- und Kommunikationstechnologie besitzen oder an einem oder mehreren gesundheitlichen Einschränkungen leiden, dennoch aber in der Lage waren, sich verständlich und präzise zu äußern.

Name	Alter	Spezifikation der Anforderungen	Besondere Bedürfnisse bei der Nutzung eines Computers
T1	73	Sehbehinderung, zitternde Hände	Anpassung der Skalierung und Bildschirmauflösung, auf Tastaturnavigation angewiesen <sup>3</sup>
T2	78	Seh- und Hörbeeinträchtigung	Anpassung der Skalierung und Bildschirmauflösung
T3	38	Blindheit	Screenreader
T4	33	Sehbeeinträchtigung (Retinitis pigmentosa)	Vergrößerungssoftware
T5	29	keine	Häufige Nutzung eines mobilen Endgerätes <sup>4</sup>
T6	27	Bruch der dominanten Hand	Einschränkung auf Tastaturnavigation <sup>5</sup>

Tab. 1: Beschreibung der Teilnehmer\*innen

<sup>3</sup> Durch zitternde Hände ist eine Maus schlecht zu bedienen

<sup>4</sup> Die Entwicklung geräteunabhängiger Web-Angebote spielt eine wichtige Rolle bei einer barrierefreien Gestaltung [YH08]. Ein geräteabhängiges UI kann digitale Barrieren beinhalten, da es nur den Endgeräten mit der benötigten Bildschirmauflösung zur Verfügung steht. Außerdem zeigt die Studie von Morris et. al., dass Menschen mit Beeinträchtigungen den Zugang zu mobilen Endgeräten haben, die den gleichen Nutzungsraten entspricht wie der allgemeinen Bevölkerung (84% der Menschen mit Beeinträchtigungen besitzen oder benutzen ein mobiles Endgerät) [Mo16]

<sup>5</sup> Aufgrund der vorübergehenden Beeinträchtigung



Das Alter dieser Gruppe reichte von 27 bis 78 Jahren (Geschlechterverhältnis 1:1). Die Teilnehmer\*innen bilden nur einen kleinen repräsentativen Durchschnitt<sup>6</sup> der Zielgruppen ab [Th10]. Eine detaillierte Beschreibung der Teilnehmer\*innen ist in der Tab.1 zu finden. Zu Beginn wurde den Teilnehmer\*innen mitgeteilt, wie die Testsitzungen ablaufen, dass sie aufgezeichnet werden und die gesammelten Informationen im Rahmen dieser Arbeit ausgewertet werden.

### **3 Prototypische Entwicklung und Evaluierung**

DT wurde in einen Problem- und Lösungsraum aufgeteilt. Diese wurden nacheinander durchgeführt. Der Problemraum bestand aus den drei DT-Phasen „Verstehen“, „Beobachten“ und „Standpunkt definieren“ und wurde in fünfmaliger Iteration durchgeführt. Im Rahmen einer umfassenden Literaturrecherche wurden Gründe nicht-barrierefreier UIs und die Herausforderungen bei der Umsetzung von Barrierefreiheit im Internet gesichtet.

Der Lösungsraum beschäftigt sich mit prototypischer Entwicklung und schließt mit einem Prototyp ab. Hier ging es darum, mit den Teilnehmer\*innen zusammen die ersten Prototypen zu erstellen, zu testen und iterativ zu verbessern. Fortführend wurde eine Testwebseite erstellt, welche bewusst viele, bekannte digitale Barrieren beinhaltete, sodass geprüft werden konnte, wie die Teilnehmer\*innen über die o.g. Webseite navigieren und mit ihr interagieren konnten. Ferner war es wichtig herauszufinden, wie es den Teilnehmer\*innen erging, wenn sie auf die Barrieren stießen und wie sie damit umgingen. Diese Informationen bildeten eine Grundlage für weitere Überlegungen, wie und mit welchen Mitteln diese Barrieren beseitigt werden konnten.

Der Lösungsraum wurde sechsmal iterativ durchgeführt und bestand aus den DT-Phasen „Testen“, „Ideen finden“ und „Beobachten“. Je Iteration wurde pro Teilnehmer\*in eine individuelle Sitzung durchgeführt (durchschnittliche Sitzungsdauer ca. 100-120 Minuten). Das Ziel jeder Iteration war es, Brainstorming durchzuführen, möglichst viele Erkenntnisse zu den Barrieren in Prototypen zu sammeln, Ideen aus der vorherigen Iteration in einen Prototyp umzusetzen und diesen dann mit den Teilnehmern zu testen. Am Ende jeder Iteration entstanden unterschiedliche Prototypen. Die einzelnen Phasen werden im Folgenden erläutert.

#### **3.1 Verstehen**

In dieser Phase ging es darum, eine Wissensbasis in Bezug auf Barrierefreiheit im Internet zu schaffen. Es wurde eine umfangreiche Literaturrecherche durchgeführt, um ein

---

<sup>6</sup> Die Barrieren im Internet sind so vielfältig und unterschiedlich wie die Behinderungen selbst. Künftig soll die Anzahl der Teilnehmerschaft erhöht werden, um den repräsentativen Durchschnitt der Zielgruppen anschaulicher darzustellen

besseres Verständnis über die Probleme der digitalen Barrierefreiheit zu erlangen, um sich dem aktuellen Forschungsstand anzunähern und sich einen breiten Überblick über das Thema „Barrierefreiheit von Web-Anwendungen“ zu verschaffen. So wurden die relevanten Zielgruppen identifiziert, die von digitalen Barrieren betroffen sind. Anschließend wurden die Bedürfnisse im Kontext der Barrierefreiheit für die wichtigsten Zielgruppen abgeleitet.

### **3.2 Beobachten**

In dieser Phase lag die Identifizierung der Bedürfnisse der Zielgruppe im Fokus. Ergänzend zu den in der Literaturrecherche gesammelten Erkenntnissen wurden im Entwicklungsprozess Interviews mit den Teilnehmer\*innen durchgeführt. Ziel der Interviews war die Beobachtung der Testpersonen und die Identifizierung konkreter Barrieren, für die Lösungen gefunden werden sollten. Die Fragen der Interviews wurden den Teilnehmer\*innen individuell gestellt und bezogen sich auf den Inhalt der jeweiligen Iteration.

### **3.3 Standpunkt definieren**

In dieser Phase wurden die gesammelten Informationen aufbereitet, interpretiert, analysiert und in einer Checkliste zusammengefasst. Die Erkenntnisse aus den vorherigen Phasen wurden zusammengefasst, um eine klare Übersicht über den aktuellen Stand dieser Arbeit zu bekommen. Anschließend wurden diese Erkenntnisse zu Clustern zusammengefasst, um ein Muster zu identifizieren. Die Checkliste wurde erstellt, um sicherzustellen, dass die gesammelten Erkenntnisse bei der Erarbeitung des Frameworks berücksichtigt wurden.

### **3.4 Ideen finden**

Diese Phase zielte darauf ab, mögliche Lösungsvorschläge durch klassisches Brainstorming zusammen mit den Teilnehmer\*innen zu erarbeiten. Ziel dieser Phase war es für die jeweilige Iteration möglichst viele unterschiedliche Ideen bezüglich der kreierte Prototypen abzuleiten. Die gewonnenen Ideen wurden dokumentiert und gruppiert.

### **3.5 Prototypen entwickeln**

In dieser Phase ging es darum für die Teilnehmer\*innen nachvollziehbare Lösungen zu entwickeln, indem Lösungsvorschläge in Form von LoFi<sup>7</sup>- und HiFi<sup>8</sup>- Prototypen für die

---

<sup>7</sup> Der Low-Fidelity-Prototyp, meist als LoFi-Prototyp bezeichnet, ist ein frühes Design eines Produkts

<sup>8</sup> Der High-Fidelity-Prototyp, auch bekannt als HiFi-Prototyp, ist ein Prototyp mit einem hohen Detailgrad an Funktionalität und Design, der dem fertigen Produkt im Aussehen ähnlich ist

entstandenen Ideen erarbeitet wurden. Die zuvor für wichtig befundenen Ideen wurden in den ersten LoFi-Prototypen umgesetzt. Durch die Prototypen waren die Teilnehmer\*innen in der Lage, die entstandenen Funktionalitäten zu verstehen, zu testen und ein Feedback zum Prototyp zu geben.

### **3.6 Testen**

Das Feedback der Teilnehmer\*innen wurde genutzt, um die Prototypen zu iterieren und diese in einem HiFi-Prototyp zu integrieren, der das Ergebnis dieses Beitrages darstellt. In dieser Phase galt grundsätzlich, dass Scheitern nicht nur erlaubt, sondern auch erwünscht ist, denn durch das Scheitern an bereits erstellten Prototypen lassen sich vorhandene Barrieren identifizieren und besser verstehen [VE12]. Die Bedürfnisse der Teilnehmer\*innen wurden in dieser Phase stärker herausgearbeitet und durch weitere Iterationen in den nächsten LoFi-Prototypen eingearbeitet.

### **3.7 Zusammenfassung**

Durch die Methode des Design Thinking konnten wichtige Erkenntnisse gewonnen werden. Diese verdeutlichten wie wichtig es ist, die tatsächliche Implementierung oder Gestaltung eines UI mit echten Benutzer\*innen durchzuführen. Der angewandte Prozess war nutzerzentriert, was dazu führte, dass Fehler während der Entwicklung schnell aufgedeckt werden konnten. Der Entwicklungsprozess war iterativ. Dies ermöglichte es, verschiedene Lösungen auszuprobieren, kontinuierlich Feedback von den Teilnehmer\*innen zu erhalten und dieses in die iterative Entwicklung des HiFi-Prototyps zu integrieren. Ein wesentlicher Nachteil des DT-Prozesses war es, dass mit den Testpersonen in mehreren Testsitzungen viel Zeit verbracht wurde. Angesichts der gesammelten Rückmeldungen aus den Iterationen war es diese Zeit jedoch wert. Der erste Schritt in dem Entwicklungsprozess war die Durchführung einer Sitzung mit den Teilnehmer\*innen. Die größte Schwierigkeit bei den ersten Sitzungen war es, den Teilnehmern\*innen grundlegendes Fachwissen zu dem Design und der Gestaltung des UIs zu vermitteln. Die Testphase war eine der wichtigsten Phasen im Entwicklungsprozess. In dieser Phase konnte mittels Storytellings das Design durchgespielt werden. Dabei konnte regelmäßig überprüft werden, ob die entwickelten Entwürfe mit den Erwartungen der Zielgruppe übereinstimmen. Im Rahmen der sechs Iterationen entstand ein Entwicklungsframework, das jede beliebige Web-Anwendung um Funktionen erweitert, die Menschen mit und ohne Behinderung die Interaktion mit einem UI erleichtert.

## **4 Entwicklungsframework für barrierefreie UIs**

Accessible User Interface (AUI) ist ein im Rahmen dieser Arbeit entstandenes, auf JavaScript basiertes Entwicklungsframework. AUI bietet Funktionen zur benutzerfreundlichen Darstellung von Webinhalten und zur Vereinfachung der

allgemeinen Navigation und Interaktion innerhalb dieser Anwendung. Der UCD-Ansatz hat gezeigt, dass AUI die Benutzerfreundlichkeit und die Barrierefreiheit einer Web-Anwendung verbessert, indem es digitale Barrieren in einem UI an vielen Stellen abbaut, Menschen mit oder ohne Beeinträchtigung die Interaktion und Navigation mit jeder Web-Anwendung erleichtert und somit einen „barriereärmeren“ Zugang schafft. Es wurden HTML und CSS-Vorlagen von Bootstrap<sup>9</sup> in dem AUI eingesetzt, um diverse Komponenten des Frameworks visuell darzustellen. JavaScript wurde eingesetzt, um die Funktionen des Frameworks zu implementieren. Nach der Einbettung des AUI-Frameworks in das HTML-Dokument und des Aufrufs des HTML-Dokuments wird der Prozess der Initialisierung angestoßen. Dabei werden die folgenden Komponenten des AUIs in das DOM<sup>10</sup> des HTML-Dokuments hinzugefügt: eine Navigation für die Bedienungshilfe, eine sekundäre Navigation zu HTML- Orientierungspunkten, Inhaltsverzeichnisse und Sprachsteuerung. Im Folgenden wird auf die einzelnen Komponenten eingegangen.

#### 4.1 Bedienungshilfe

Nach der Einbettung und Aktivierung des AUI wird im HTML-Dokument ein Button für die Bedienungshilfe (s. Abb. 1) an einer fixierten Position (unten links) platziert, wodurch er auch beim Scrollen dauerhaft eingeblendet bleibt. Durch den Button wird ein Besucher über das Thema Barrierefreiheit im Internet aufmerksam gemacht.



Abb. 1: Button für die Bedienungshilfe

Durch das Betätigen des Buttons wird die Navigation für die Bedienungshilfe eingeblendet (s. Abb. 2). Hier hat man die Möglichkeit zwischen, den Komponenten „Personalisierte Darstellung“, „Inhaltsverzeichnis“, „Tastenkombinationen“, „Information“ und „Schließen“ zu wählen.



Abb. 2: Navigation für die Bedienungshilfe

**Personalisierte Darstellung** - Die personalisierte Darstellung stellt eine der wesentlichen Stärken des AUI dar. Diese ermöglicht eine übersichtliche Darstellung vom Inhalt einer Webseite. Dabei werden störende Elemente wie die Seitennavigation, Schaltflächen,

<sup>9</sup> <https://getbootstrap.com/docs/3.3/getting-started/> (Stand: 15.08.2019)

<sup>10</sup> Document Object Model (DOM) - Darstellung der HTML-Dokumente als eine Baumstruktur

Werbung und Hintergrundbilder etc. ausgeblendet, sodass die Seitenbesucher\*innen sich auf den eigentlichen Inhalt der Webseite konzentrieren kann. Die Einstellungen der „Personalisierten Darstellung“ können den persönlichen Vorstellungen entsprechend angepasst werden. Sie ermöglicht es, die Schriftgrößen, den Zeilen- sowie Buchstabenabstand anzupassen, die Schriftart zu ändern oder das Layout zu wechseln, um dadurch den Kontrast der Farben zu ändern. Diese Einstellungen können zu jeder Zeit zurückgesetzt werden. Durch das Betätigen des Buttons „Zurück zur Navigation“ wird die Leseansicht deaktiviert und die Web-Anwendung in ursprünglicher Form dargestellt.

**Tastenkombination** - Durch Betätigen des Navigationspunktes „Tastenkombination“ erhält man eine Liste der Tastenkürzel, die man als Bedienungshilfe einsetzen kann. Diese Tastenkombinationen sind bei den gängigen Browsern nahezu gleich und ermöglichen es, z.B. die Darstellung der aktuellen Web-Anwendung zu vergrößern oder zu verkleinern oder an den Anfang oder ans Ende der Web-Anwendung zu springen. Tastaturkürzel sind hilfreich, wenn die Anwender\*innen keine Maus zur Hand haben. Außerdem bieten sie einen zusätzlichen Nutzen für motorisch eingeschränkte Personen oder Personen, deren Sehvermögen eingeschränkt ist.

**Information** - Durch Betätigen des Navigationspunktes „Information“ erhalten die Seitenbesucher\*innen Hinweise zur allgemeinen Orientierung und Informationen darüber, wie die einzelnen Komponenten der Bedienungshilfe (z.B. personalisierte Darstellung, Sprachsteuerung, etc.) zu bedienen sind.

## 4.2 Inhaltsverzeichnis

Bei langen Beiträgen kann man als Nutzer\*innen den Überblick schnell verlieren. Ein übersichtliches Inhaltsverzeichnis hilft den Seitenbesucher\*innen dabei, direkt auf eine gewünschte Textstelle zu springen. Auch Sehbehinderte und Blinde können davon profitieren, da sie nicht mehr den ganzen Beitrag durchsuchen müssen und über diese Funktion direkt zur gewünschten Stelle gelangen. AUI generiert eine Navigation, durch die man sich ein Inhaltsverzeichnis für alle Überschriften und eins für alle Verlinkungen anzeigen lassen kann.

## 4.3 Sekundäre Navigation zu HTML- Orientierungspunkten

Bei der Initialisierung wird am Anfang des HTML-Dokuments eine sekundäre Navigation (siehe Abb. 3) hinzugefügt. Diese ist für sehende Besucher\*innen ausgeblendet und erscheint erst, wenn ein/e Anwender\*in über die Tabulatortaste navigiert. Somit ist diese Navigation für die Seitenbesucher\*innen, die über eine Mause navigieren, nicht wahrnehmbar. Die Navigation beinhaltet die folgenden vier Sprungmarken. Diese sind wie folgt beschrieben:

- „Weiter mit Bedienungshilfe“: Nach dem Betätigen dieser Sprungmarke wird die Navigation für die Bedienungshilfe aufgerufen.

- „Zur Navigation“: Nach dem Betätigen dieser Sprungmarke wird zur Seitennavigation gesprungen.
- „Zum Inhaltsverzeichnis“: Dieses Element ruft die oben beschriebene Inhaltsverzeichnisse auf.
- „Zum Inhalt“: Hierdurch springt man zum Hauptinhalt der Web-Anwendung gesprungen.



Abb. 3: Sekundäre Navigation zu HTML- Orientierungspunkte

Alle Komponenten diese Navigation lassen sich über AUI konfigurieren und anpassen.

#### 4.4 Sprachsteuerung

Mittels Sprachsteuerung sind Anwender\*innen in der Lage, durch die Stimme Befehle an das AUI-Framework zu übermitteln, welche das Steuerelement aktiviert bzw. auf welche Verlinkung ein Klick-Ereignis ausgelöst werden soll. Die Sprachsteuerung wurde mittels „Web Speech API“<sup>11</sup> erstellt. Web Speech API ist ein funktionierender Prototyp und bislang nur in Chrome und Firefox<sup>12</sup> verfügbar. Durch das zweimalige Betätigen der Steuerungstaste (Strg, Ctrl) wird die Sprachsteuerung aktiviert (s. Abb. 4). Durch die Aktivierung wird der HTML-Baum (DOM) des aufgerufenen Dokuments durchsucht. Dabei werden alle Verlinkungen und Schaltflächen identifiziert und mit einem eindeutigen Schlüssel (ID) versehen. Diese werden in der Form eines Shortcuts für die Anwender\*innen visuell dargestellt. Die Anwender\*innen können nun die ID desjenigen Elements aussprechen, das er aufrufen will. Wird hingegen die Audioaufnahme erfasst, so wird auf dem gewählten Element ein Klick-Ereignis aufgelöst.

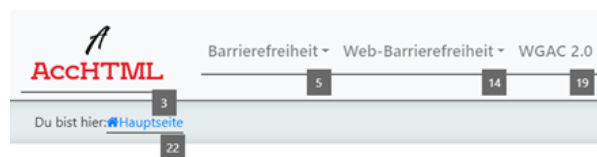


Abb. 4: Aktivierte Sprachsteuerung in einem UI

<sup>11</sup> <https://wicg.github.io/speech-api/> (Stand: 22.09.2019)

<sup>12</sup> Chrome und Firefox sind die Webbrowser

## 5 Fazit

Bei der Durchführung dieser Arbeit lag der Schwerpunkt nicht nur auf der Erarbeitung eines UI-Frameworks für barrierearme UI durch DT, sondern auch auf der Erprobung der Methode, von der Einführung in die Methodik über die Ideengenerierung bis hin zum Prototyping und Test. Die Entscheidung, ein Framework für die Web-Anwendungen zu erarbeiten, hat sich bewährt. Dieses war nach einer ausführlichen Analyse-Phase sehr aufwendig und mit viel Arbeit verbunden. Da die Entwicklung parallel mit den Iterationen erfolgte, mussten zudem oft die neuen Erkenntnisse an verschiedenen Stellen nachgetragen werden. In den Iterationen konnte miterlebt werden, wie es den betroffenen Personen erging, wenn sie auf eine Barriere stießen. Dies waren besonders beeindruckende Erfahrungen. Hierbei wurde das Thema digitale Barrierefreiheit aus der Sicht von blinden, sehbehinderten, vorübergehend beeinträchtigten und älteren Personen betrachtet. Diese stellen nur einen kleinen Kreis aller Menschen mit Beeinträchtigung oder Behinderungen dar. Diese Arbeit ist ein erster großer Schritt in Richtung des Abbaus von digitalen Barrieren und der Anpassung eines UIs an individuelle Bedürfnisse. Das AUI-Framework soll durch weitere Endnutzer\*innen getestet werden, um weitere Barrieren aufzuzeigen und diese zu beheben. Durch die Brainstorming-Sitzungen mit den Teilnehmer\*innen entstanden mögliche Lösungsvorschläge, die in den HiFi-Prototypen mit einem hohen Detailgrad an Funktionalität umgesetzt wurden. Diese wurden dann mit den Teilnehmer\*innen getestet, bewertet und anschließend verbessert. Das Thema „Digitale Barrierefreiheit“ betrifft auch die Autoren, die Inhalte einer Web-Anwendung erstellen. Deren Aufgabe ist es, Inhalte auf eine verständliche Weise darzustellen, denn eine Webanwendung ist nur dann barrierefrei, wenn sowohl ihre Struktur und Funktionalitäten als auch ihr Inhalt barrierefrei sind.

## 6 Ausblick

Digitale Barrierefreiheit wird in Zukunft aufgrund ihrer verpflichtenden Umsetzung eine immer größere Rolle im Bildungswesen spielen. Der Fortschritt der Technik wird sowohl einen positiven als auch negativen Einfluss auf diesem Gebiet haben. Es werden immer neue mobile internetfähige Geräte entwickelt, welche neue Barrieren mit sich bringen. Laut Lazar et al. [La13] beeinflusst die Unwissenheit der Webentwickler\*innen und der Führungskräfte die Barrierefreiheit der Web-Angebote erheblich. AUI zeigt, dass es möglich ist, eine mit Barrieren belastete Web-Anwendung mittels eines UI-Frameworks barriereärmer zu machen. Jedoch gibt es noch viele Barrieren, die den Zugang zu den Informationen oder E-Learning Angebot unnötig erschweren. Das AUI-Framework bietet für Menschen mit kognitiven oder neurologischen Einschränkungen sowie mit Hör-, Lernbeeinträchtigungen noch keine Hilfestellung. In weiteren Studien muss untersucht werden, wie man diese Personengruppen durch das AUI unterstützen kann. Die Erhebung des aktuellen Standes der Umsetzung der Barrierefreiheit im Internet erfordert ebenfalls weitere Studien.

## Literaturverzeichnis

- [Dr14] Dreifus, Florian; Simon, Katarina: Design thinking - neuer impuls für das business process reengineering, Tech. Report 2, Institut für Projektmanagement und Informationsmodellierung (IPIM), 2014.
- [HP11] Hellbusch, Jan Eric; Probiesch, Kerstin: Barrierefreiheit verstehen und umsetzen, Webstandards für ein zugängliches und nutzbares Internet, dpunkt verlag, Heidelberg, 2011.
- [He14] Henry, Shawn Lawton; Shadi Abou-Zahra; Judy Brewer: The Role of Accessibility in a Universal Web, Web for All Conference on W4A, Seoul Republic of Korea, 2014.
- [LO09] Loiacono, E. T. et.al.: The state of corporate website accessibility. Communications of the ACM/52(9), 128-132,2009.
- [La13] Lazar, J. et.al.: A longitudinal study of state government homepage accessibility in Maryland and the role of web page templates for improving accessibility. Government Information Quarterly/30, 289–299, 2013.
- [Me07] Meron, Michael: Barrierefreiheit im World Wide Web: Analysen zu Bedarf und Umsetzbarkeit, Hamburg, 2007.
- [Mo16] Morris, J. et.al.: Smartphone Use and Activities by People with Disabilities: User Survey 2016. Journal on Technology and Persons with Disabilities. Santiago, J. (Eds): CSUN Assistive Technology Conference, 51-66, 2017.
- [Sc17] Schallmo, Daniel R.A.: Design Thinking erfolgreich anwenden: So entwickeln Sie in 7 Phasen kundenorientierte Produkte und Dienstleistungen. Berlin Heidelberg New York: Springer-Verlag, 2017.
- [Th10] Thesmann, Stephan: Einführung in das Design multimedialer Webanwendungen. Berlin Heidelberg New York: Springer-Verlag, 2010.
- [VE12] Vetterli, Christophe; et.al.: Die Innovationsmethode Design Thinking. Dynamisches IT-Management: So steigern Sie die Agilität, Flexibilität und Innovationskraft Ihrer IT. Düsseldorf: Symposion Publishing, 289-310, 2012.
- [Ve18] Velleman, Eric Martin: The implementation of web accessibility standards by dutch municipalities, factors of resistance and support, 2018.
- [WC19] W3C, Über das W3C, [https:// www.w3c.de/about/](https://www.w3c.de/about/), Stand: 12.08.2019.



## Ermittlung von Usability-Kriterien für Jugendliche mit kognitiven Beeinträchtigungen im Bereich E-Learning

### Durchführung einer ethnografischen Feldforschung

Arlett Semm<sup>1</sup> und Ulrike Spierling<sup>2</sup>

**Abstract:** Die zunehmende Digitalisierung im Schulbereich wirft die Frage auf, welche Usability-Anforderungen an Computersysteme, Internetinhalte und E-Learning-Plattformen für Jugendliche mit kognitiven Beeinträchtigungen bestehen. Mit Hilfe einer über ein Schuljahr dauernden ethnografischen Feldforschung mit teilnehmender Beobachtung in einer Förderschule für geistige Entwicklung entstand eine umfangreiche Sammlung aus Beobachtungsprotokollen und Transkripten. Im Mittelpunkt der Beobachtungen stand der Unterricht mit seinen Methoden und Routinen. Ein Schwerpunkt lag auf der Erfassung der Art und Weise wie Wissen vermittelt und aufgenommen wird. Die Auswertung ergab einen komplexen Usability-Anforderungskatalog, dessen Beachtung bei der Konzeption von Lernmedien für diese Zielgruppe deren Teilhabe erleichtern bzw. ermöglichen kann.

**Keywords:** Usability, E-Learning, Anforderungen, Kriterien, kognitive Beeinträchtigung, Lernschwierigkeiten

## 1 Einleitung

Kenntnisse im Bereich der Informations- und Kommunikationstechnik (IKT) speziell im E-Learning sind zu einem integralen Bestandteil der Bildung geworden. Besonders die Beherrschung digitaler Lernmedien spielt hierbei eine wesentliche Rolle. Die Nutzung von Online-Lern-Systemen zum Selbstlernen hat in den letzten Jahren immer mehr zugenommen. In Zeiten von Pandemien wie durch CoVid-19 ist die digitale Entwicklung noch einmal beschleunigt worden. Neue Technologien werden zunehmend in Schule, Ausbildung, Hochschule und Beruf eingeführt. Deshalb ist es für alle wichtig, Zugang zu digitalen Inhalten zu bekommen, um nicht aus der Informations- und Bildungsgesellschaft ausgeschlossen zu werden. Im Bildungsbereich sind die Chancen für Menschen mit Beeinträchtigungen geringer. Der Anteil der Beeinträchtigten geht von der Grundschule bis zur Hochschule deutlich zurück. Bei Menschen mit kognitiven Beeinträchtigungen ist die bestehende Integration in das schulische Bildungssystem noch geringer, da diese Art der Beeinträchtigung als stark lernbehindernd gilt. Online Lernplattformen

---

<sup>1</sup> Ernst-Abbe-Hochschule Jena, Fachbereich Wirtschaftsingenieurwesen, Carl-Zeiss-Promenade 2, 07743 Jena, Arlett.Semm@eah-jena.de

<sup>2</sup> Hochschule RheinMain, Fachbereich Design Informatik Medien, Unter den Eichen 5, 65195 Wiesbaden, ulrike.spierling@hs-rm.de

könnten helfen, diese Situation zu verbessern. Sie haben den Vorteil, dass sie jederzeit verfügbar sind und sie für jede(n) die Möglichkeit bieten, im eigenen Tempo mit beliebigen Wiederholungen zu lernen. Allerdings gibt es einen Mangel an Forschung und Veröffentlichungen über die Nutzung von und den Zugang zu IKT wie E-Learning für Menschen mit kognitiven Beeinträchtigungen, insbesondere im Vergleich zu mehr körperlicheren Behinderungen [CGS19] [De15]. Um sich mit dieser Zielgruppe zu beschäftigen, sollte zunächst geklärt werden, was für diese spezielle Gruppe zu beachten ist. Oftmals liegt bei den Betroffenen eine Kombination aus verschiedenen Beeinträchtigungen vor, was die zielgruppengerechte Konzeption einer Software für sie noch schwieriger macht. Studien zu diesem Thema beschäftigen sich meist mit spezifischen Syndromen oder Störungen, wie z.B. der Autismus-Spektrum-Störung oder dem Down-Syndrom. Die derzeitige Forschung sollte sich laut Cinquin et. al. jedoch nicht auf die Art der Beeinträchtigung konzentrieren, sondern auf die Fähigkeiten, welche eingeschränkt sind, wie z.B. das Gedächtnis, die Lesefähigkeit oder die Aufmerksamkeit [CGS19]. Hierbei ist es wichtig zu erforschen, welche Kriterien erfüllt sein müssen, um die Nutzung digitaler Lernmedien für diese vielschichtige Zielgruppe zu ermöglichen.

Allgemeine Anforderungen an Nutzbarkeit sind in der Usability-Norm ISO 9241-11 (Gebrauchstauglichkeit) beschrieben [DI18], und Richtlinien für barrierefreie Webinhalte des World Wide Web Consortiums (W3C) liegen in Form der Web Content Accessibility Guidelines (WCAG) 2.1 vor [We18]. Darüber hinaus gibt es Gesetze wie das deutsche Behindertengleichstellungsgesetz (BGG) [Bu20a] und Festlegungen in Form der Barrierefreien Informationstechnik-Verordnung (BITV) [Bu20b] und der UN-Behindertenrechtskonvention [Be17]. Bei der Softwareentwicklung wird großer Wert auf die Kenntnis von Benutzer\*innen und ihrer Umgebung gelegt, daher wird erhoben, welche Ziele sowie Aufgaben mit der Software oder Webseite erreicht werden sollen. Allerdings gibt es nur wenige Erfahrungen mit kognitiv beeinträchtigten Menschen als Nutzer\*innen. Das Problem ist, dass verantwortliche Entwickler\*innen wenig Verständnis für die speziellen Bedürfnisse der Zielgruppe haben, zuweilen auch Berührungsängste [De15]. Gleichzeitig fällt es vielen kognitiv beeinträchtigten Menschen nicht leicht, sich selbst oder ihre Wünsche und Bedürfnisse auszudrücken, und viele Methoden, Routinen und Verhaltensweisen werden eher unbewusst im Alltag angewendet. Ein Zugang zu dieser Thematik wird in multidisziplinären Langzeitstudien in der realen Umgebung mit den Endbenutzer\*innen, Sonderpädagog\*innen und Forscher\*innen gesehen. Auf diese Weise können Lösungen gefunden werden, die sicherstellen, dass kognitiv beeinträchtigte Benutzer\*innen digitale Lernmedien selbstständig nutzen können [CGS19] [De15]. Eine Form der Langzeitstudie ist die ethnographische Feldforschung. Dabei wird eine bestimmte Gruppe oder Organisation vor Ort für einen längeren Zeitraum in ihrem Alltag beobachtet und an diesem teilgenommen. Ein längerer Zeitraum ist notwendig, um sich mit der anfangs fremd erscheinenden Umgebung und den Menschen vertraut zu machen und Einzelheiten wahrzunehmen, für die vorab keine Fragen vorbereitet waren. Ein besonderer Schwerpunkt liegt dabei auf der Beobachtung der sozialen und organisatorischen Interaktionen im Nutzerumfeld mit einer gewissen Fremdheit. Der Vorteil der Ethnographie ist die Erfassung sehr reichhaltiger Daten über den Nutzerarbeitskontext.

Andererseits hat dies zur Folge, dass große Datenmengen (Mitschriften, Audioaufzeichnungen) ausgewertet werden müssen. Zudem kann nur eine begrenzte Nutzergruppe betrachtet werden, obwohl die Zeiträume der Studie sehr lang sind [Ro08]. Dem gegenüber steht die immer schneller werdende Softwareentwicklung, weshalb diese Form der Studie im IKT-Umfeld kaum genutzt wird.

## 2 Methodik

Die Motivation für die hier beschriebene Arbeit bestand darin, die Wissenslücke über spezifische Anforderungen der genannten Zielgruppe an digitale Lernmedien zu verringern. Konkret ging es darum, Usability-Kriterien für Computer, Internet und E-Learning für Jugendliche mit besonderen Lernschwierigkeiten im Lernumfeld zu ermitteln. Um einen tieferen Einblick in die speziellen Bedürfnisse der Zielgruppe zu erhalten, die diese selbst schlecht artikulieren kann, wurde aus den oben genannten Gründen eine ethnographische Feldforschung durchgeführt. Zu diesem Zweck folgte die Erstautorin dieses Beitrags ein Jahr lang zweimal wöchentlich als teilnehmende Beobachterin dem Unterricht einer Klasse der Oberstufe in einer Förderschule mit dem Förderschwerpunkt „geistige Entwicklung“.

Folgende Forschungsfragen wurden im Rahmen der ethnografischen Feldforschung untersucht:

- Wie wird im Unterricht Wissen vermittelt und wie wird es aufgenommen?
- Welche Methoden werden im Unterricht (mit und ohne Computer) angewendet?
- Welche Routinen werden zum Lernen, zur Hilfe und zur Lernkontrolle eingesetzt?
- Welche Probleme und Anforderungen haben die Lehrer\*innen?

Die Gruppe in der besuchten Klasse bestand aus 4 Jungen, 5 Mädchen (Jugendliche im Alter von 14 bis 18 Jahren), einer Lehrerin und einem Erzieher. Die Auswahl der Kohorte ergab sich durch die Unterstützung der betreuenden Lehrerin, die zur notwendigen Zustimmung der Schule und der Eltern führte. Alle Schüler\*innen hatten individuelle und spezielle Bedürfnisse, Fähigkeiten und Interessen. Einige Schüler\*innen hatten keinerlei motorische Einschränkungen, während andere große Probleme in der Feinmotorik oder sogar in der Grobmotorik hatten. Manche konnten sehr gut lesen und schreiben, andere beherrschten dies noch nicht, oder hatten Probleme, Texte sinnerfassend zu lesen. Häufig auftretende Probleme waren Aufmerksamkeits- und Konzentrationsdefizite; diese behindern die Lernstoffvermittlung. Die Konfrontation und der Umgang mit Fehlern bereiteten vielen der Schüler\*innen Probleme. Meistens konnten Inhalte in Textform in Kombination mit passenden Bildern besser verstanden werden. Viele Schüler\*innen hatten bereits zu Hause erste Erfahrungen mit dem Computer und dem Internet gesammelt. Die Vielfalt dieser Schülergruppe war für ihre Lehrer\*innen eine besondere Herausforderung. Um eine optimale Lernumgebung für jede(n) einzelne(n) Schüler\*in

zu schaffen, mussten die Lehrkräfte individuell auf jede(n) eingehen können. Ein weiteres Problem war, dass einige Lehrer\*innen und Erzieher\*innen selbst nur geringe Erfahrungen mit dem Computer und dem Internet hatten. In der besuchten Klasse jedoch wurden der Computer und das Internet regelmäßig zur Lösung von Aufgaben im Unterricht eingesetzt. Für die Studie wurde der Unterricht auf Tonband aufgezeichnet, die wichtigsten Momente wurden transkribiert. Diese ausführlichen Transkriptionen enthalten die Originalaussagen der Schüler\*innen und der Lehrer\*innen. Darüber hinaus wurden die Beobachtungen, Erkenntnisse und Ideen in den Beobachtungsprotokollen auf insgesamt über 350 Seiten festgehalten: 11 Beobachtungsprotokolle ohne Transkriptionen und 24 Beobachtungsprotokolle mit Transkriptionen [Se10].

Während der Erhebungsphase wurde bereits mit der sequenziellen Auswertung der Beobachtungsprotokolle und Transkripte begonnen. Es fand dabei ein ständiger kritischer Austausch mit den Sonderpädagog\*innen über die Ergebnisse statt. Die ermittelten Kriterien wurden mit Hilfe der Beobachtungsprotokolle weiter verifiziert. Im Anschluss wurden die Kriterien kategorisiert und gruppiert.

### 3 Ausgewählte Ergebnisse

Eines der Ergebnisse der ethnografischen Feldforschung und der anschließenden sequenziellen Auswertung war eine komplexe Mindmap, bestehend aus 115 strukturellen und inhaltlichen Usability-Kriterien für E-Learning, Computer- und Internetanwendungen für kognitiv beeinträchtigte Jugendliche. Der Übergang zwischen diesen Kategorien ist allerdings nicht trennscharf (vgl. Grauschattierungen in Tabelle 1). Die einzelnen Kriterien traten im Verlauf der Gesamtbeobachtung mit unterschiedlicher Häufigkeit auf. Im Rahmen dieses Beitrages wird im Folgenden eine Auswahl an Ergebnissen hervorgehoben, welche besonders relevant für die Zielgruppe erscheinen (vgl. Tab. 1). Weitere umfangreiche Daten sind in der zugrundeliegenden Masterarbeit [Se10] zu finden.

Nr.	Kriterium
20.	Audioausgabe der Texte
40.	kurze Texte mit beschreibenden eindeutigen Bildern (Inhalt vermitteln)
58.	Inhalte mit Beispielen stützen
60.	bei Beschreibung Alltagsbezüge nutzen
74.	Mitdenken ermöglichen (Brainstorming)
89.	Gruppenarbeit ermöglichen (Kooperation)
107.	keine Mehrbelastung für die Lehrer*innen

Tab. 1: Auszug aus dem Kriterienkatalog: Zusammenspiel struktureller (weiß) und inhaltlicher Kriterien (dunkelgrau); sowie gemischter Kriterien (hellgraue Felder).

Alle in Tab. 1 ausgewählten Kriterien adressieren in einem komplexen Zusammenspiel das Problemfeld des erleichterten Textverstehens. Ist z.B. eine Audioausgabe von Texten (Kriterium 20) auch ein Kriterium bei anderen Arten der Beeinträchtigung (wie Sehvermögen), ergibt sich die Motivation hier aus der Beobachtung, dass im Unterricht üblich war, alle Textinhalte auszusprechen. So war wenig Lesen erforderlich, um alle, auch Schüler\*innen mit Leseschwierigkeiten, in das Thema einzubeziehen. Das Kriterium 60, bei Beschreibungen Alltagsbezüge zu nutzen, und das Kriterium 89, in Gruppen zu arbeiten, sind typische Routinen, welche mindestens 30-mal während der teilnehmenden Beobachtung auftraten. Ein Beispiel eines typischen Alltagsbezugs ist die Erklärung zum Verständnis und zur Einordnung der Uhrzeit „20:15 Uhr“, als „die Zeit, zu der die Filme beginnen“. Die Übernahme des wiederholten individuellen Vorlesens durch eine Audioausgabe, wahlweise unterstützt durch Kopfhörer, entlastet Lehrer\*innen zeitlich, um andere individuelle Unterstützung zu leisten.

Die Notwendigkeit für individuelles Mitdenken und -sprechen in Wiederholungen zeigte sich unter anderem am Beispiel eines besser lesenden Schülers aus der teilnehmenden Beobachtung: Schüler I liest auf einer Webseite im Internet laut, monoton und fließend vor. Schüler I: „Lehrerin1 ich hab alles durchgelesen. (Beobachterin: „Und was besagt das?“) [Schüler I atmet schwer belastet.] Also, naja, da steht eben in einiges beschrieben, dass das so und so gemacht wurde und ich hab mir das alles durchgelesen. Das ist sehr viel [...]“ [..] Beobachterin: „er hat jetzt zwar gut gelesen, aber“ (Lehrerin1: „der kann das besser übers hören[...]“). Der Text war mit 30 Sätzen und knapp über 300 Wörtern für den Schüler relativ lang, aber in leichter Sprache verfasst. Schüler I hat diesen Text auf der Webseite gelesen. Das Lesen allein hat ihn so beschäftigt, dass er den Inhalt nicht aufnehmen konnte. Lehrerin1 erwähnte, dass ihm dies besser mittels Hören gelingt. In der Fachliteratur wird die zusätzliche Audioausgabe von Texten für Nutzer\*innen mit Leseschwierigkeiten als mögliches Kriterium genannt [HB05], welches je nach Bedarf von diesen verwendet werden kann, sonst bleiben laut Hellbusch vielen Nutzer\*innen wichtige Inhalte verschlossen. Die Inhalte können entweder in natürlicher oder synthetischer Sprache angeboten werden [HB05].

#### **4 Zusammenfassung und Ausblick**

Durch die diesem Beitrag zu Grunde liegenden Untersuchungen wurden nicht nur Usability-Kriterien gesammelt und kategorisiert, sondern insbesondere für einige dieser Kriterien ein vertieftes Verständnis der komplexen Nutzungssituation erreicht, was hier am Beispiel der Audioausgabe demonstriert wurde. Innerhalb der durchgeführten Studie wurden einige der ermittelten Kriterien noch einmal kritisch hinterfragt, indem sie in prototypisch selbst erstellte digitale Lernelemente testweise integriert wurden, um Nutzertests durchführen zu können. Den Besonderheiten der Zielgruppe, die nicht mit traditionellen Methoden zu befragen ist, wurde mit einem ethnografischen Ansatz Rechnung getragen. Empathie mit den Benutzergruppen (Schüler\*innen und Lehrer\*innen) sowie mit Nutzungssituationen wurde dabei priorisiert gegenüber einem Anspruch auf Voll-

ständigkeit hinsichtlich sämtlicher nötiger Usability-Anforderungen. Weitere Untersuchungen, insbesondere mit weiteren Prototypentwicklungen und Tests, sind daher nötig, um den Kriterienkatalog zu vervollständigen und die auf speziell kognitive Einschränkungen ausgerichteten Kriterien abzugrenzen von anderen Anforderungen an Barrierefreiheit.

## Literaturverzeichnis

- [Be17] Beauftragte der Bundesregierung für die Belange von Menschen mit Behinderungen: Die UN-Behindertenrechtskonvention. Übereinkommen über die Rechte von Menschen mit Behinderungen. [https://www.behindertenbeauftragte.de/SharedDocs/Publicationen/UN\\_Konvention\\_deutsch.pdf](https://www.behindertenbeauftragte.de/SharedDocs/Publicationen/UN_Konvention_deutsch.pdf), 2017. Stand: 22.06.2020.
- [Bu20a] Bundesamt für Justiz: Gesetz zur Gleichstellung von Menschen mit Behinderungen, <http://www.gesetze-im-internet.de/bgg/>, Stand: 22.06.2020.
- [Bu20b] Bundesamt für Justiz: 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](https://www.gesetze-im-internet.de/bitv_2_0/BJNR184300011.html), Stand: 22.06.2020.
- [CGS19] Cinquin, P.-A.; Guitton, P.; Sauzéon, H.: Online e-learning and cognitive disabilities: A systematic review. *Computers and Education*, Elsevier 130/19, S. 152-167, 2019.
- [De15] Dekelver, J.; Siplivaya, M.; Shabalina, O.; Borblik, J.; Pidoprigora, A.; Romanenko, R.: Design of Mobile Applications for People with Intellectual Disabilities. In (Kraevets, A.; Shcherbakov, M.; Kultsova, M.; Shabalina, O. Hrsg.): *Creativity in Intelligent Technologies and Data Science*. Communications in Computer and Information Science, vol 535, Springer, Cham, S. 823-836, 2015.
- [DI18] DIN-Normenausschuss Ergonomie (NAErg): DIN EN ISO 9241-11 Ergonomie der Mensch-System-Interaktion - Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte (ISO 9241-11:2018), <https://www.din.de/de/mitwirken/normenausschuesse/naerg/normen/wdc-beuth:din21:279590417>, 2018. Stand: 28.03.2020.
- [HB05] Hellbusch, J. E.; Bühler, C.(Hrsg.): *Barrierefreies Webdesign. Praxishandbuch für Webgestaltung und grafische Programmoberflächen*, 1. Auflage, dpunkt verlag, Heidelberg, 2005.
- [Ro08] Rosenthal, G.: *Interpretative Sozialforschung. Eine Einführung*, Juventa Verlag, Weinheim und München, 2008.
- [Se10] Semm, A.: *Umsetzung ermittelter Kriterien für Computer und Internet anhand eines Online Lehr- und Lerntools für Jugendliche mit erschwerten Lernprozessen. Prototyp einer Online-Lernplattform*. Masterthesis, Fachhochschule Erfurt, 2010.
- [We18] *Web Content Accessibility Guidelines (WCAG) 2.1*, [www.w3.org/TR/WCAG21/](http://www.w3.org/TR/WCAG21/), 2018. Stand: 28.03.2020.

## Kompetenzmodellierung für die grundlegende Programmierausbildung

### Eine kritische Diskussion zu Vorzügen und Anwendbarkeit der Anderson Krathwohl Taxonomie im Vergleich zum Kompetenzmodell der GI

Natalie Kiesler<sup>1</sup>

**Abstract:** Dieser Forschungsbeitrag untersucht Kompetenzanforderungen an Informatik-Studierende in der grundlegenden Programmierausbildung. Basierend auf einer qualitativen Inhaltsanalyse aktueller Lehr- und Lernziele deutscher Hochschulen und der anhand von Interviews erhobenen Perspektive von Hochschullehrenden werden angestrebte Programmierkompetenzen der Basisausbildung identifiziert. Da das Kompetenzmodell der Gesellschaft für Informatik diverse Defizite aufweist, werden die entwickelten inhaltlichen Kategorien in die Anderson Krathwohl Taxonomie (AKT) kognitiver Bildungsziele eingeordnet. Daraufhin erfolgt eine Überarbeitung der Dimensionen und Subtypen der AKT hin zu einem für die Informatik spezifischen Modell mit dem Ziel, Programmierkompetenzen gemäß ihrer kognitiven Komplexität klassifizieren zu können. Die präsentierte Handreichung ist auf den Ebenen Lehrveranstaltungskonzeption sowie Entwicklung und Bewertung von Assessments zur Messung von Programmierkompetenzen anwendbar.

**Keywords:** Programmierkompetenz, Wissensdimensionen, kognitive Prozessdimensionen

## 1 Einleitung

Der Entwicklung kontextspezifischer Kompetenzmodelle [Kl04, We01] wurde in der Informatik bisher wenig Beachtung geschenkt. Zum einen ist die Informatik eine sich stets weiterentwickelnde Fachdisziplin, zum anderen benötigen valide Instrumente zur Kompetenzmessung empirische Daten als Basis [Ko08]. Aktuell werden Kompetenzmodelle für die universitäre Informatik-Ausbildung nur in vereinzelten Forschungsarbeiten berücksichtigt [Li13, KTB16, Sc12]. Durch die Gesellschaft für Informatik (GI) wurde ein Kompetenzmodell als Empfehlung für die Entwicklung von Kompetenzen im Informatik-Studium vorgestellt [Ge16]. Obwohl deren Modell auf der Anderson Krathwohl Taxonomie (AKT) [AK01] basiert, werden einige Wissensdimensionen sowie kognitive Prozessdimensionen nicht berücksichtigt und bezüglich Kontextspezifika der Informatik fehlinterpretiert. Die vorliegende Forschungsarbeit untersucht daher die Anwendbarkeit der AKT [AK01] als anerkannten Rahmen zur Kategorisierung kognitiver Kompetenzen in der grundlegenden Programmierausbildung - dem Kern eines jeden Informatik-Studiums.

---

<sup>1</sup> Goethe-Universität Frankfurt, Akademie für Bildungsforschung und Lehrerbildung, Juridicum, Senckenberganlage 31-33, 60325 Frankfurt am Main, kiesler@em.uni-frankfurt.de

## 2 Kompetenzmodelle in der Informatik

Ziel des von der GI [Ge16] entwickelten, an die AKT angelehnten Modells ist es, die in Informatik-Studiengängen angestrebten kognitiven Kompetenzen zu beschreiben um Hilfestellung zur Gestaltung, Weiterentwicklung und Beurteilung derselben zu ermöglichen [Ge16]. Das Modell differenziert in der ersten von zwei Zeilen vier kognitive Prozessdimensionen: *Verstehen*, *Anwenden*, *Analysieren* und *Erzeugen*. Durch die beiden Zeilen soll zwischen *Typen des Wissenschaftlichen Arbeitens* mit den Subtypen T1 bis T6 unterschieden werden, die durch *Stufen der Kontextualisierung von Anwendungen* (K1 bis K5) und *Wissensdimensionen* (W1 bis W4) charakterisiert werden. Erläuterungen zur Bedeutung der Wissensdimensionen für die Informatik und zur reduzierten Darstellung in Form von zwei Zeilen fehlen jedoch. Implizierte Annotationen zu den jeweiligen Typen T, K und W in den Beispielen der 17 Inhaltsbereiche bleiben aus. Definierte Kategorien (K5) bleiben ungenutzt und die Entwicklung metakognitiver Kompetenzen scheint ausschließlich der Promotion vorbehalten zu sein. Indem die Klassifizierungen in den Beispielen von den Autoren selbst nicht präzise und nachvollziehbar angewendet werden können, offenbart sich eine der größten Schwächen des Modells. Lediglich kognitive Prozessdimensionen werden in den Beispielen erkennbar (wenn auch nicht nachvollziehbar) zugeordnet.

Die zweite Zeile des Modells beinhaltet weitere Stufen kognitiver Prozessdimensionen: 2a *Übertragen* und 3a *Bewerten*. Die aus der AKT bekannte Dimension *Erinnern* wird wegen vermeintlicher Redundanz nicht berücksichtigt. Die Felder in der zweiten Zeile unter Stufe 1 und 4 bleiben leer, weil „sich die Prozessdimension ‚Verstehen‘ eher auf grundlegende Zusammenhänge beziehen soll und Stufe 4 ‚Erzeugen‘ im Bachelor-Studiengang kaum grundlegende wissenschaftliche Innovation in komplexeren Anwendungszusammenhängen erwarten lässt“ [Ge16, S. 10]. Demnach wird die Dimension *Erzeugen* in Bachelor-Studiengängen nicht angestrebt und bleibt ungenutzt. Spätestens dadurch wird die Fehleinschätzung zum kognitiven Anspruch allein der Programmierung deutlich.

Gegenläufig zum GI-Modell liefert die AKT bereits ein komplexes Kontinuum kognitiver Kompetenzen und Wissensdimensionen. Eine weitere, explizite Differenzierung nach Kontextualisierung von Inhalten durch die GI ist nicht notwendig, da die Wissensdimensionen kontextspezifisches Wissen anhand der Subtypen adressieren. Darüber hinaus fasst das GI-Modell an anderer Stelle Dimensionen zusammen, die zu differenzierende kognitive Prozesse abbilden, wie zum Beispiel *Erinnern* und *Verstehen*. In der Informatik und speziell der Programmierung müssen Studierende stets grundlegende Fakten und Informationen aus dem Langzeitgedächtnis abrufen, so z. B. Schlüsselworte, Operatoren, Wertebereiche, etc. Das Erinnern an derartige Details ist ohne jegliches Verständnis möglich, sodass die Position der GI weiter konterkariert wird. Besonders scharfe Kritik bedarf die Position zur Kategorie des Erzeugens. Während die GI damit einhergehende kognitive Prozesse Bachelor- und Master-Arbeiten vorbehält, vertritt die Autorin die Position, dass bereits das Schreiben kleiner Algorithmen und Programme planerische, gestalterische und konstruktive Leistungen zur Produktion neuer Lösungen erforderlich macht. Diese kognitiven Prozesse sind entsprechend der AKT in die Dimension des *Erzeugens* einzuordnen.



Zusätzlich werden metakognitive Kompetenzen im Bachelor-Studium benötigt, um Lernprozesse zu organisieren, systematisch Probleme zu lösen und Selbsterkenntnis zu erreichen. Laut GI werden diese Kompetenzen erst ab der Promotion ausgebildet [Ge16].

Da die Einordnungen der GI entlang der verbleibenden kognitiven Prozessdimensionen nicht den Typen, Subtypen und Kategorien kognitiver Prozesse entsprechen, bedarf es einer Handreichung für die Klassifizierung von Kompetenzen entlang der AKT-Matrix in der Informatik. Besonders die Programmierung kann von einem komplexeren Modell in Anlehnung an die AKT profitieren, um die Gestaltung von Lehrangeboten für NovizInnen zu optimieren, da relevante kognitive Prozesse und Wissensdimensionen darin abgebildet werden. So kann zum einen das Bewusstsein von Lehrenden für den kognitiven Anspruch der Programmierung geschärft, zum anderen können Aufgaben entlang der verschiedenen Dimensionen entwickelt und damit lernförderlicher gestaltet werden. Zukünftige (Self-) Assessments könnten Programmierkompetenzen damit differenzierter abbilden.

### 3 Forschungsdesign und Ergebnisse

Die Anderson Krathwohl Taxonomie zur Klassifikation von Bildungszielen (AKT) bildet kognitive Prozesse und Wissensdimensionen in ausreichender Komplexität unter Berücksichtigung von Fachspezifika ab. Eine Reduktion in Kombination mit neuen Dimensionen zur Kontextualisierung, wie im GI-Modell gezeigt, erscheint paradox. Aufgrund der Fehlinterpretationen der GI wird die AKT vor dem Hintergrund der kognitiven Anforderungen in der Programmierung als Kern jedes Informatik-Studiums neu bewertet. Folgende Forschungsfragen werden untersucht: (1) *Inwieweit kann die Anderson Krathwohl Taxonomie in der Grundlagenausbildung der Programmierung Anwendung finden, um Programmierkompetenzen zu klassifizieren?* (2) *Wie fachspezifisch ist die AKT zu begreifen?*

Zunächst wurde eine qualitative Inhaltsanalyse aktueller Kompetenzziele deutscher Hochschulen im Bereich der grundlegenden Programmierausbildung durchgeführt. Ergänzend wurde anhand von Experten-Interviews die Perspektive von Hochschullehrenden erfasst. Die zusammengefassten und abstrahierten inhaltlichen Kategorien von Kompetenzen wurden in die Dimensionen AKT eingeordnet. Daraufhin erfolgte eine Überarbeitung der Dimensionen und Subtypen der AKT hin zu einem für die Informatik spezifischen Modell.<sup>2</sup>

Als Ergebnis zeigt Abb. 1 die neu interpretierten Wissensdimensionen der AKT mitsamt Subtypen für die Basisausbildung der Programmierung. Darüber hinaus weist Abb. 2 die entsprechende Bedeutung der kognitiven Prozessdimensionen aus, indem Definitionen und Beispiele präsentiert werden. Die beiden Tabellen sind als kontextspezifische Adaption der AKT für die grundlegende Programmierung zu verstehen und können als Handreichung zur Klassifikation von Kompetenzzielen dieser Domäne genutzt werden.

---

<sup>2</sup> Die Klassifizierung aller kognitiven Kompetenzen sowie die vollständige Übersicht nicht-kognitiver Kompetenzen wurde bereits veröffentlicht [Ki20]. Die Analyseergebnisse sind anhand des Codebuchs bzw. der Übersicht codierter Segmente online verfügbar (<https://github.com/nkiesler-cs/delfi2020.git>).

TYPEN UND SUBTYPEN	DEFINITION UND BEISPIELE FÜR DIE PROGRAMMIERUNG
<b>A. FAKTENWISSEN – Grund</b>	legende Inhalte und Details, die Studierende für die Einarbeitung in ein Thema wissen müssen
AA. Wissen um Terminologie	Fachvokabular (z. B. Sprachelement wie Schlüsselwörter, Escape-Sequenzen, Literale, binäres Zahlensystem, Operatoren, elementare Datentypen, Ausnahmen einer Programmiersprache)
AB. Wissen um spezifische Details und Elemente	Fachspezifische Ressourcen und Quellen (z. B. Geschichte der Informatik, des Computers, der Programmiersprachen, Phasen der Softwareentwicklung; Kenntnis zuverlässiger Lehrbücher und Werkzeuge zum Erlernen des Programmierens)
<b>B. KONZEPTIONELLES WISSEN – Übergeordnete Zusammenhänge zw. Elementen als Teile einer übergeordneten Struktur</b>	
BA. Wissen um Klassifikationen und Kategorien	Wissen um Programmierparadigmen, Klassen der Zeitkomplexität und algorithmischen Effizienz, Eigenschaften von Algorithmen, Generationen von Programmiersprachen, numerische Datentypen, Chomsky-Hierarchie als Klassifikation für formale Sprachen
BB. Wissen um Prinzipien und Verallgemeinerungen	Kenntnisse der Prinzipien von Programmierparadigmen und Programmiersprachen (z. B. Typsysteme, Namensgebung, Funktionen, Parameterübergabe, Speicherverwaltung)
BC. Wissen um Theorien, Modelle und Strukturen	Kenntnis der Komplexitätstheorie (Effizienz, Laufzeitkomplexität), Modelle von Computerarchitekturen, Kenntnis von Speichermodellen und Rechnermodellen (z. B. Turingmaschine)
<b>C. PROZEDURALES WISSEN – Methoden und Regeln für die Anwendung von Algorithmen, Techniken, Strategien</b>	
CA. Wissen um fachspezifische Strategien und Algorithmen	Syntax und Semantik von Programmiersprachen, Algorithmen für typische Probleme kennen, Code schreiben, Umwandeln von Dezimalzahlen in Fließkommazahlen (oder anderes Zahlensystem)
CB. Wissen um fachspezifische Techniken und Methoden	Konventionen für guten Programmierstil, Methodik zur Analyse von Problemen und Algorithmen, Werkzeuge zur Softwareentwicklung kennen und nutzen, Modellierung und Entwurf von Algorithmen, Algorithmen testen und debuggen
CC. Wissen um Kriterien zur Bestimmung der Eignung von Verfahren	Angemessene Verwendung eines Entwurfsmusters oder Algorithmus zur Lösung eines unbekanntem Problems (z. B. Auswahl eines problemadäquaten Entwurfsmusters oder Sortierverfahrens)
<b>D. META-KOGNITIVES WISSEN – Allgemeines Wissen über Kognition sowie Bewusstsein und Kenntnis der eigenen Kognition</b>	
DA. Strategisches Wissen	Wissen um Organisation von Lernprozessen und wann welche Lernstrategien für welchen Zweck sinnvoll eingesetzt werden sollten
DB. Wissen über kognitive Aufgaben	Wissen über Bedingungen kognitiver Aufgaben (z. B. nicht-lösbare Probleme, wann zusätzliche Ressourcen zur Problemlösung nötig werden, wann Möglichkeiten zum Transfer von Wissen auf neue Aufgaben bestehen, Bewusstsein über systematische Ansätze zur Problemanalyse und -lösung)
DC. Selbsterkenntnis	Reflexion persönlicher Stärken und Schwächen, Bewusstsein über verwendete Strategien, Angemessenheit des Selbstvertrauens und der Selbstwirksamkeit, Bewusstsein über Ziele, Motivation und persönliche Interessen, Übernahme von Verantwortung für Lernprozesse

Abb. 1: Wissensdimensionen in der Programmierausbildung

## 4 Diskussion

Die Einordnung der Programmierkompetenzen in die AKT macht den hohen kognitiven Anspruch an StudienanfängerInnen der Informatik im Rahmen der Programmierausbildung deutlich. In kaum einem anderen Studienfach (mit Ausnahme weniger anderer Natur- bzw. Ingenieurwissenschaften) werden von Novizen in derartigem Umfang konstruktive Leistungen erwartet. In der Programmierung stellt das Schreiben eigenen Codes als Methode zur Problemlösung eine konstruktive Leistung dar. Paradoerweise wird in den analysierten Daten nur in geringem Ausmaß Faktenwissen und Konzeptionelles Wissen auf der Ebene des Erinnerns und Verstehens ausgezeichnet. Gleichzeitig häufen sich Kompetenzziele in der Dimension *Erzeugen* von *Prozeduralem Wissen*. Die explizite Ausformulierung niedrigerer kognitiver Prozesse dieser Wissensdimension wurde häufig ausgespart. Wenngleich diese Aussparungen anteilig aufgrund des inkludierenden Charakters der höheren Dimensionen zu erwarten waren, sollten Ziele in niedrigeren kognitiven Prozessdimensionen explizit formuliert werden. Die fehlende Sichtbarmachung der weniger

KOGNITIVE PROZESSE UND ALTERNATIVEN	DEFINITION UND BEISPIELE FÜR DIE PROGRAMMIERUNG
<b>1. ERINNERN</b> – Aufrufen von	von bekannten Informationen aus dem Langzeitgedächtnis
1.1 ERKENNEN Identifizieren	Zugreifen auf Wissen des Langzeitgedächtnisses, das mit vorgelegtem Material konsistent ist (z. B. Daten wichtiger Ereignisse kennen: wann der Computer erfunden wurde und von wem)
1.2 ERINNERN Abrufen	Abrufen von relevantem Wissen aus dem Langzeitgedächtnis (z. B. Erinnern an Daten wichtiger Ereignisse in der Geschichte der Informatik)
<b>2. VERSTEHEN</b> – Ableiten	von Bedeutung aus Instruktion und Kommunikation
2.1 INTERPRETIEREN Erklären, Umschreiben, Darstellen, Übersetzen	Wechseln von einer Darstellungsform (numerisch, schriftlich) zu einer anderen (verbal) (z. B. paraphrasieren von Literatur, Dokumentationen, Handbüchern; schriftliche Aufgabenstellung mündlich beschreiben)
2.2 VERDEUTLICHEN Illustrieren, Instanzieren	Aufzeigen eines spezifischen Beispiels bzw. Illustration eines Konzepts oder Prinzips (z. B. Stack als Beispiel eines abstrakten Datentyps erläutern)
2.3 KLASSIFIZIEREN Kategorisieren, Subsumieren	Zugehörigkeit zu einer Kategorie oder Klasse bestimmen (z. B. zu einer Klasse von Datentypen; Integer als primitiven Datentyp klassifizieren)
2.4 ZUSAMMENFASSEN Abstrahieren, Generalisieren	Zusammenfassen eines allgemeinen Themengebiets oder Motivs, bzw. eines oder mehrerer wichtiger Punkte (z. B. Kernkonzepte der jeweiligen Programmierparadigmen zusammenfassen)
2.5 ENTNEHMEN Schlussfolgern, Extra-/Interpolieren, Vorhersagen	Aus präsentierten Informationen logische Schlussfolgerungen ziehen (z. B. Zustände von zwei Booleschen Variablen logisch verknüpfen und Wahrheitswert ableiten)
2.6 VERGLEICHEN Gegenüberstellen, Abgleichen, Zuordnen	Erkennen von Übereinstimmungen zwischen zwei Ideen, Objekten, Konzepten (z. B. Gegenüberstellen von Datentypen, Datenstrukturen oder Algorithmen; Von-Neumann und Harvard-Architektur vergleichen)
2.7 ERKLÄREN Konstruieren von Modellen	Konstruktion eines Modells zu Ursache/Wirkung eines Systems (z. B. unterschiedlichen Speicherplatzverbrauch und Laufzeit von Iteration vs. Rekursion erklären; Von-Neumann-Architektur erklären)
<b>3. ANWENDEN</b> – situative	Ausführung, Anwendung oder Implementierung von Prozessen, Vorgängen oder Strategien
3.1 AUSFÜHREN Durchführen	Anwenden eines Verfahrens auf eine bekannte Aufgabe (z. B. Dezimalzahlen in Dualzahlen, Oktalzahlen, Hexadezimalzahlen umkodieren)
3.2 UMSETZEN Verwenden	Anwenden eines Verfahrens auf eine unbekannte Aufgabe (z. B. Qualitätskriterien und Programmierkonventionen auf eigenen Quellcode anwenden)
<b>4. ANALYSIEREN</b> – Zerlegung	von Material in Teile und logische Erklärung durch Klarlegung derer Beziehung, Bedeutung im Gesamtkontext, Organisation oder Charakterisierung
4.1 DIFFERENZIEREN Unterscheiden, Orientieren, Auswählen	Unterscheiden zwischen relevanten und irrelevanten bzw. wichtigen und unwichtigen Teilen eines präsentierten Materials (z. B. eine Problemstellung zerlegen und zur Problemlösung relevante Aspekte auswählen)
4.2 ORGANISIEREN Zusammenhänge finden, integrieren, skizzieren, zerlegen, strukturieren	Bestimmen, wie Elemente innerhalb einer Struktur zusammenpassen oder funktionieren (z. B. Bestandteile fremden Codes zerlegen und dadurch dessen Funktionsweise skizzieren).
<b>5. BEWERTEN</b> – begründete	Auswahl aus mehreren Alternativen, Beurteilung durch Überprüfung von Kriterien und Standards
5.1 PRÜFEN Koordinieren, Entdecken, Überwachen, Testen	Erkennen von Inkonsistenzen in Prozessen oder Produkten; Erkennen der Wirksamkeit eines Verfahrens während der Umsetzung (z. B. Testen von Algorithmen und Programmen auf Korrektheit und Eigenschaften, Fehlersuche in Programmen, Verantwortung übernehmen für Lernerfolg)
5.2 KRITISIEREN Beurteilen	Erkennen von Inkonsistenzen zwischen einem Produkt und extern formulierten Kriterien oder Standards; Erkennen positiver und negativer Eigenschaften eines Produkts; Erkennen der Angemessenheit eines Verfahrens für ein Problem (z. B. Beurteilen, welcher Algorithmus bzw. welches Programm das Angemessenste ist, um ein bestimmtes Problem zu lösen; Komplexität von Algorithmen beurteilen)
<b>6. ERZEUGEN</b> – Zusammenstellung	von Elementen und Einzelteilen zu einem neuen, funktionierenden Ganzen, (Einzigartigkeit, Originalität) durch Planung, mentale (Re-)Strukturierung, Konzeption, Produktion
6.1 GENERIEREN Hypothesen bilden	Probleme auf neue Art darstellen und dadurch alternative Hypothesen und Möglichkeiten zur Problemlösung aufstellen; Grenzen von Vorwissen und Theorien werden überschritten (z. B. Modellieren von Problemen; Verschiedene neue Algorithmen zur Lösung eines Problems zusammentragen (Iteration vs. Rekursion); Transfer von Erkenntnissen auf neue Probleme und deren Lösung)
6.2 PLANEN Designen	Bewusste oder unbewusste Ausarbeitung eines geeigneten Verfahrens zur Erfüllung einer Aufgabe, ggf. Festlegen von Teilzielen und Arbeitsschritten (z. B. Modellieren eines Programms; Algorithmenentwurf; Datenstrukturen entwerfen; Schnittstellen entwerfen)
6.3 PRODUZIEREN Konstruieren	Probleme auf Arbeitsplan zur Problemlösung verfolgen und damit ein Produkt erfinden oder entwickeln, das den gestellten Anforderungen gerecht wird (z. B. Programmiersprachliche, lauffähige Lösungen für Probleme schreiben; GUIs programmieren; Systematik zur Problemlösung entwickeln)

Abb. 2: Kognitive Prozessdimensionen in der Programmierausbildung

komplexen kognitiven Kompetenzen könnte eine Ursache für zu hohe Erwartungen an ProgrammieranfängerInnen sein, die von McCracken et al. [Mc01] identifiziert wurden.

Weiterhin wird durch den hohen Abstraktionsgrad der AKT eine Gültigkeit für alle Fachgebiete erreicht. Jedoch können leicht missverständliche Interpretationen erfolgen, wie das angelehnte GI-Modell zeigt. Die erschwerte Zuordnung der einzelnen Dimensionen ist vor allem der Ambiguität der Schlüsselworte geschuldet. Einige vermeintliche Beispiele, wie etwa „Konzept“, können sogar in die Irre führen, da sie eben nicht zwangsläufig auf *Konzeptionelles Wissen* verweisen. Daher ist eine fachspezifische Variation der AKT, insbesondere für die Informatik unabdingbar. Wie das adaptierte Beispiel in Tabelle 1 und 2 zeigt, kann die AKT in der Grundlagenausbildung der Programmierung durchaus Anwendung finden, um Programmierkompetenzen zu klassifizieren. Die allgemeinen und stark abstrahierten Dimensionen bedürfen lediglich einer fachspezifischen Adaption um auch in der Informatik Anwendung zu finden, z. Bsp. bei der Überarbeitung von Curricula oder der Entwicklung von (adaptiven) Assessments.

## Literaturverzeichnis

- [AK01] Anderson, L. W.; Krathwohl, D. R.: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. 2001.
- [Ge16] Gesellschaft für Informatik e.V.: Empfehlungen für Bachelor und Masterprogramme im Studienfach Informatik an Hochschulen. 2016.
- [Ki20] Kiesler, N.: Towards a Competence Model for the Novice Programmer Using Bloom's Revised Taxonomy. In: Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education. S. 459-465, 2020.
- [KI04] Klieme, E.: Was sind Kompetenzen und wie lassen sie sich messen? Pädagogik, 56(6):10-13, 2004.
- [Ko08] Koeppen, K. et al.: Current issues in competence modeling and assessment. Zeitschrift für Psychologie/Journal of Psychology, 216(2):61-73, 2008.
- [KTB16] Kramer, M.; Tobinski, D. A.; Brinda, T.: Modelling Competency in the Field of OOP: From Investigating Computer Science Curricula to Developing Test Items. In: Stakeholders and Information Technology in Education. S. 1-8, 2016.
- [Li13] Linck, B. et al.: Competence model for informatics modelling and system comprehension. In: IEEE Global Engineering Education Conference. S. 85-93, March 2013.
- [Mc01] McCracken, M. et al.: A Multi-National, Multi-Institutional Study of Assessment of Programming Skills of First-Year CS Students. In: Working Group Reports from ITiCSE. S. 125-180, 2001.
- [Sc12] Schäfer, A. et al.: The Empirically Refined Competence Structure Model for Embedded Micro- and Nanosystems. In: Proceedings of the 17th ACM Annual Conference on Innovation and Technology in Computer Science Education. S. 57-62, 2012.
- [We01] Weinert, F.: Concept of competence: A conceptual clarification. In: Defining and selecting key competencies. 2001.

## Erwerb von Digitalen Kompetenzen für die Lebens- und Arbeitswelt

Tobias Falke <sup>1</sup>

**Abstract:** In der modernen Wissensgesellschaft haben Informations- und Kommunikationstechnologien (IKT) einen bedeutenden Stellenwert eingenommen. Seit vielen Jahren beschäftigt sich der Autor mit Modellen von Medienkompetenz und deren Erwerb an unterschiedliche Zielgruppen. In der aktuellen bildungspolitischen Debatte wird oft der Begriff der digitalen Kompetenz genutzt. Dies macht deutlich, dass der Inhalt von Medienkompetenz sich gewandelt hat und heute mit digitaler Kompetenz gleichgesetzt werden kann. Ziel des hier vorgestellten Vorhabens war es, ein Veranstaltungskonzept zum Erwerb digitaler Kompetenz zu entwickeln, welches auf der inhaltlichen Ebene von Lehrenden der Brandenburgischen Technischen Universität (BTU) Cottbus-Senftenberg getragen werden kann. Dazu startete im Oktober 2019 eine Ringvorlesung, die das aktuelle Thema „Digitalisierung in der Gesellschaft“ fokussiert. Die Vorlesungsreihe richtete sich an die allgemeine interessierte Öffentlichkeit außerhalb des Campus und an Studierende die im Rahmen des fachübergreifenden Studiums (FÜS) die Veranstaltung besucht haben. Dieser Beitrag beschreibt die Entwicklung des Konzepts, abschließend werden Evaluationsergebnisse vorgestellt.

**Keywords:** Blended Learning, digitale Kompetenzen, Medienkompetenz, digitaler Wandel, online Lehre

### 1 Ausgangslage

Der digitale Wandel ist Teil unserer Lebenswirklichkeit. Wir befinden uns in einem tiefgreifenden Transformationsprozess, der unsere Art zu kommunizieren, zu lernen, zu wirtschaften und zu arbeiten verändert. „[...] die fortschreitende Digitalisierung ist zum festen Bestandteil unserer Lebens-, Berufs- und Arbeitswelt geworden.“ „Die Länder beziehen in ihre Lehr- und Bildungspläne sowie Rahmenpläne [...] die Kompetenzen ein, die für eine aktive, selbstbestimmte Teilhabe in einer digitalen Welt erforderlich sind.“ [Km17]

### 2 Zielsetzung

Die Brandenburgische Technische Universität Cottbus-Senftenberg (BTU) hat sich dieser Herausforderung gestellt und im Rahmen des Qualitätspakt Projektes „Exzellenz

---

<sup>1</sup> BTU Cottbus-Senftenberg, Multimediazentrum/IKMZ, Platz der Deutschen Einheit 2, 03046 Cottbus, tobias.falke@b-tu.de

von Studium und Lehre - Individueller Studieneinstieg, Innovative Studienmodelle, Forschendes Lernen“ ein Konzept für eine Lehrveranstaltung entwickelt, welches das Thema „Digitale Kompetenzen für die Lebens- und Arbeitswelt“ in den Fokus rückt. Dem Anspruch gerecht zu werden, dass dieses Thema auch für die regionale Bürgerschaft relevant und von Interesse ist, wurde beschlossen eine Ringvorlesung zu entwickeln, die gleichzeitig auch als Fachübergreifendes Studium (FüS) durchgeführt werden kann. Um die Leistungsansprüchen von 6 Credit Points (CP) an ein FÜS-Modul an der BTU zu erfüllen, wurde diese Veranstaltung im Blended Learning Format durchgeführt. Dazu haben die beteiligten Dozenten sieben Online Phasen entwickelt.

### **3 Theoretische Vorbetrachtungen**

Erste Überlegungen zu einem geeigneten Format zum Erwerb Digitaler Kompetenzen wurden im interdisziplinären Team des Multimediazentrums im Oktober 2018 vollzogen. Im Rahmen von Gruppendiskussionen mit studentischen und wissenschaftlichen Hilfskräften und mit Teamkollegen wurden aus didaktischer und technologischer Sicht relevante Schlagworte und Themen für eine Veranstaltung für Studierende zum Thema digitale Kompetenz in einem diskursiven Prozess, erarbeitet.

Ausgangspunkt weiterer Entwicklung waren verschiedene Modelle zum Erwerb digitaler Kompetenz. Die große Gesellschaftsstudie D21-Digital-Index bildet die gesamte deutsche Wohnbevölkerung ab. Um mit den Aussagen der Studie die Anschlussfähigkeit an Europa zu gewährleisten, wurde die Digitalkompetenzen des D21-Digital-Index mit denen des Europasses 4 [Eu18] abgeglichen und an ihm orientiert. Beide Erhebungsinstrumente, welche auf Selbsteinschätzung basieren, zeigen im Ergebnis fünf Bereiche notwendiger Digitalkompetenzen auf: Datenverarbeitung/ Information; Kommunikation; Sicherheit; Problemlösung und Erstellung von Inhalten [In19].

Um ein Kompetenzmodell mit dem im Rahmen der zu entwickelnden Veranstaltung gearbeitet wird, zu definieren, wurden weitere Modelle betrachtet. Der DIGCOMP-Framework der EU [FPB13], [Vu16], [GVP17], [RP17] hat für dieses Vorhaben das Problem, das der Kompetenzbegriff sehr weit gefasst wird und eher auf den „mündigen Bürger“ abzielt. Weitere Kompetenzmodelle wie das Technological Pedagogical Content Knowledge (TPCK) - Modell [MK06] das Modell Digital Bildung [So03], [KJ13] wurden betrachtet und analysiert, hier standen Lehrende an Schulen im Fokus der Analysen. Als weitere Grundlage für die Entwicklung des notwendigen Kompetenzrasters diente darum das Digital-Literacy-Framework des Joint Information Systems Committee (JISC.) Diese britische Organisation unterstützt Universitäten und Colleges mit digitaler Infrastruktur und berät akademische Einrichtungen beim Einsatz digitaler Technologien [Ji12], [Ji14]. Das Digital-Literacy-Framework adressiert Lehrende an akademischen Einrichtungen und wird bereits von verschiedenen Hochschulen im angelsächsischen Raum eingesetzt. Das Modell des JISC diente Eichhorn [Ei19] als Ausgangspunkt für die Weiterentwicklung. Das Modell zeigt die acht Dimensionen: IT-Kompetenz, Informationskompetenz, Kommunikations-/Kollaborations-kompetenz, Digitale Lehre, Digitale Identität und

Karriereplanung, Digitale Wissenschaft, Digital Produzieren, Analyse-/Reflektionskompetenz, auf, wobei in drei Kompetenzstufen differenziert wird. Stufe 1: Überblickswissen / Grundlagen digitaler Kompetenzen; Stufe 2: Praktische Anwendung im Lehr-/Lernkontext bzw. der eigenen Forschungstätigkeit; Stufe 3: Anleitung und Begleitung von Studierenden und/oder Kolleginnen und Kollegen.

Zielgruppe der Veranstaltung sind Studierende aller Studiengänge und aller Fachsemester, aber auch im Rahmen der Ringvorlesung, interessierte Bürger. Als Ausgangspunkt für weitere Überlegungen für die zu entwickelnden Inhalte wurde sich für die fünf Bereiche notwendiger Digitalkompetenzen des D21-Digital-Index [In19] und für die Systematik nach Eichhorn et.al. [Ei19] entschieden. Die von Eichhorn et.al. dargestellten Kompetenzstufen finden insofern Beachtung, dass sich die Organisatoren vorgenommen haben, Kompetenzen im Rahmen der Ringvorlesung lediglich auf Stufe 1 zu entwickeln, da die Veranstaltung lediglich den Anspruch hat, dem Teilnehmer ein Überblickswissen und Grundkompetenzen im Themenfeld Digitalisierung zu vermitteln. Die Studierenden des FÜS-Moduls hingegen sollen aufgrund der zu bearbeitenden sieben Online-Phasen auch Stufe 2: Die praktische Anwendung im Lehr-/ Lernkontext, erreichen.

#### 4 Umsetzung

Um die acht Bereiche notwendiger Digitalkompetenzen im Rahmen der Veranstaltung zu bearbeiten, wurden mögliche Inhalte und Themen abgeleitet und Lernziele erarbeitet. Zwischen den einzelnen Veranstaltungen der Vorlesung wurden orts- und zeitunabhängige bearbeitbare Online-Phasen durchgeführt, die mit der Vergabe von Punkten bewertet werden. Die Online-Phasen bestehen aus (Selbst-) reflexionsaufgaben, thematischen Vertiefungen und Gruppenaufgaben, die in schriftlicher Form ausgearbeitet wurden. Die Erstellung der Online- Aufgaben wurde kooperativ mit den jeweiligen Referenten oder deren Mitarbeitenden durchgeführt. Anspruch war es, die Aufgaben sehr vielfältig, multimedial und multimodal für die teilnehmenden Studierenden zu gestalten.

Nr.	Titel der Veranstaltung	Onlineaufgabe
1	Lehren und Lernen im digitalen Zeitalter	10 Online-Aufgaben im Themenfeld
2	Data Literacy: Digitale Kompetenzen in der Hochschulbildung	Keine Onlinephase
3	Sollten Arbeitnehmer neue Technologien annehmen oder sich ihnen widersetzen?	Textaufgabe mit dem Ergebnis Argumente zusammenzufassen und ein Fazit zu ziehen
4	Cottbus – ein gallisches Dorf in Sachen Digitalisierung	Keine Onlinephase
5	Digitale Prozesse in Produktion und Logistik	Entwicklung einer App – Digitales Fahrtenbuch für den B2B Bereich

		Business-to-Business (Abkürzungen B2B, B-to-B oder BtB) bezeichnen Geschäftsbeziehungen zwischen zwei oder mehr Unternehmen)
6	Digitale Entwurfs- und Fabrikationsmethoden	Internetrecherche zu Anwendungsgebieten der Digitalisierung im Baugewerbe
7	Belastungen durch Technikstress erfolgreich begegnen	Internetrecherchen und Fragestellungen zu Texten zu den Themen: Wandel der Arbeitswelt und Technikstress
8	Digitale Selbstausbeutung oder gesunder Umgang mit Neuen Medien?	Erarbeitung eines Mediennutzungsprotokolls, Reflexion und Diskussion in der Gruppe zu Mediennutzungsgewohnheiten
9	Zensur, Anonymität und Meinungsmanipulation im Internet	Diskussion per Videokonferenz zu den Themen: Manipulationsmaschine - Soziale Netzwerk - Fake News
10	Datenschutz im Marketing und in der Unternehmensorganisation	Keine Onlinephase

Tab. 1: Vortragsthemen und Inhalte der Onlinephase

## 5 Gestaltung der Onlinephasen

Die Studierenden absolvieren nach jeder von BTU-Lehrenden angebotenen Vorlesung eine Onlinephase zur Festigung und Vertiefung des vermittelten Vorlesungsstoffes. Nach Gilly Salmon vollziehen sich virtuelle Gruppenprozesse in fünf Phasen, die durch Aktivitäten (E-tivity's) unterstützt werden [Sa02]. „E-tivities“, sind nach Salomon der Schlüsselfaktor für erfolgreiches Online-Lernen und zielen darauf ab, gemeinsame Online-Aktivitäten zu initiieren und den Reflexionsprozess zu fördern. Mit "E-tivity" bezeichnet Salmon "Aufträge für aktives und interaktives Online-Lernen".

## 6 Evaluation und Auswertung

Am FÜS Modul „Digitale Kompetenzen in der Lebens- und Arbeitswelt“ haben im Wintersemester 2019/20 38 Studierende teilgenommen, 32 Studierende haben das Modul erfolgreich abgeschlossen. Im Rahmen der Analyse der Kursdaten wurden 32.744 Zugriffe erfasst, somit wurde im Durchschnitt 9,13 mal pro Tag und Teilnehmer\*innen auf den Kurs zugegriffen. In Summe wurden 5.393 Beiträge von den Studierenden verfasst, durchschnittlich sind dies somit 1,5 Beiträge pro Teilnehmer pro Tag.

Da uns die Weiterentwicklung des Moduls sehr wichtig ist, hat das Referat Qualitätsmanagement in Studium und Lehre eine auf qualitativen Methoden beruhende



Veranstaltungsevaluation durchgeführt. Die Studierenden haben sich auf Basis der Modulbeschreibung für das FÜS Modul entschieden. Dabei waren die inhaltlichen Themen und die zu entwickelnden Kompetenzen und Arbeitsweisen mit technischen Hilfsmitteln gleichermaßen ausschlaggebend. Zudem machte das neue didaktische Konzept der Verbindung von Präsenz- und Online- Anteilen die Studierenden neugierig.

Die ganze Gruppe war sich einig, dass sich der Mut der Verantwortlichen in jedem Fall gelohnt hat, ein neues Format mit Online- und Präsenzteilen zu konzipieren und umzusetzen. Gelobt wurde zudem die Themenvielfalt bzw. die Verbindung der verschiedenen Fachgebiete. Die Vorträge wurden als hilfreiches Inputformat bewertet und sollten als Präsenzformat unbedingt beibehalten werden. Wichtig ist den Studierenden, dass diese im Nachgang als Videoaufzeichnung zur Verfügung gestellt werden. Die Vorteile der flexiblen Bearbeitung von Inhalten zeigten sich für die Studierenden in Form wertgeschätzter Autonomie, z.B. bei der eigenständigen Quellenrecherche zur Lösung der Aufgaben.

## 7 Ausblick

Die Evaluation und Auswertung hat gezeigt, dass das Format weiterentwickelt werden sollte. Die Verzahnung aus digitalen und analogen Angeboten hat sich für dieses Veranstaltungsformat als sehr gewinnbringend herausgestellt und wurde von den Studierenden für den „Lehrpreis der BTU“ vorgeschlagen. Damit die Inhalte einer noch breiteren Öffentlichkeit zur Verfügung gestellt werden können, ist es das Ziel, in einem zu beantragenden Fortsetzungsprojekt, die entstandenen Vorlesungsaufzeichnungen, Begleitmaterialien und Bearbeitungsaufgaben in einem Online-Kurs zusammenzuführen.



## Literaturverzeichnis

- [Ei19] Eichhorn, Michael (2019). Fit für die digitale Hochschule? – Modellierung und Erfassung digitaler Kompetenzen von Hochschullehrenden, MedienPädagogik Nr. 36. Unter: <https://www.medienpaed.com/article/download/710/677/> am 14.03.2020
- [Eu18] Europäische Union, Europasses 4 - Digitale Kompetenzen – Raster zur Selbstbeurteilung, 2018, [https://euopass.cedefop.europa.eu/sites/default/files/dc\\_-\\_de.pdf](https://euopass.cedefop.europa.eu/sites/default/files/dc_-_de.pdf) am 11.03.2020
- [FPB13] Ferrari, Anusca; Punie, Yves; Brečko, Barbara: DIGCOMP – A framework for developing and understanding digital competence in Europe, Scientific and technical, research Publications Office, Luxembourg, series 26035, 2013
- [Gi20] Gilroy, Patrick: Fähigkeiten für die digitale Welt – Engagement als Chance. ZiviZ im Stifterverband, Berlin, 2020, <http://www.ziviz.de/download/file/656> am 11.03.2020
- [GVP17] Gomez, Shamilia C.; Vuorikari, Rina; Punie, Yves: DigComp 2.1. – The Digital Competence Framework for Citizens with eight proficiency levels and examples of use, Scientific and technical research series, Publications Office, Luxembourg, 2017

- [In19] Initiative D21 Wie digital ist Deutschland - Eine Studie der Initiative D21, Kantar, 2019, [https://initiated21.de/app/uploads/2020/02/d21\\_index2019\\_2020.pdf](https://initiated21.de/app/uploads/2020/02/d21_index2019_2020.pdf) am 16.03.2020
- [Ji12] Developing Digital Literacies: Briefing Paper, 2012, [http://www.jisc.ac.uk/media/documents/publications/briefingpaper/2012/Developing\\_Digital\\_Literacies.pdf](http://www.jisc.ac.uk/media/documents/publications/briefingpaper/2012/Developing_Digital_Literacies.pdf) am 16.03.2020
- [Ji14] Developing Digital Literacies: Overview, 2014, <https://www.jisc.ac.uk/guides/developing-digital-literacies> am 16.03.2020
- [Km17] KMK (2017). Bildung in der digitalen Welt – Strategie der Kultusministerkonferenz; [https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie\\_2017\\_mit\\_Weiterbildung.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2018/Digitalstrategie_2017_mit_Weiterbildung.pdf) am 16.03.2020
- [KJ13] Krumsvik, Rune J., Jones, Lise O.: Teachers Digital Competence in Upper Secondary School: 2013; <http://www.icitc.org/Proceedings2013/Papers%202013/05-1-Krumsvik.pdf> am 11.03.2020
- [MK06] Mishra, Punya & Koehler, Matthew J.: Technological Pedagogical Content Knowledge: A new framework for teacher knowledge. Teachers College Record 108, 2006, <https://www.punyamishra.com/wp-content/uploads/2008/01/mishra-koehler-tcr2006.pdf> am 12.03.2020
- [RP17] Redecker, Christine; Punie, Yves: European Framework for the Digital Competence of Educators: Dig CompEdu, Publications Office of the European Union, Luxembourg, 2017, [http://publications.jrc.ec.europa.eu/repository/bitstream/JRC107466/pdf\\_digcomedu\\_a4\\_final.pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/JRC107466/pdf_digcomedu_a4_final.pdf) am 11.03.2020
- [Sa00] Salmon, Gilly: E-moderating: the key to teaching and learning online, Kogan Page Limited, London, 2000
- [Sa02] Salmon, Gilly: E-tivities: The key to active online learning, Kogan Page Limited, London, 2002
- [So03] Soby, Morton: Digital Competence: from ICT skills to digital “Bildung”, University of Oslo, Oslo, 2003
- [Vu16] Vuorikari, Riina; Punie, Yves; Carretero, Santiago; Brande, Lievi Van den: DigComp 2.0: The Digital Competence Framework for Citizens. Update Phase 1, Publications Office. European Commission, Luxembourg, 2016 <http://dx.publications.europa.eu/10.2791/11517> am 11.03.202

## Medienkompetenz fördern und Medieneinsatz befördern

### Didaktische Überlegungen bei der Entwicklung einer mobilen Anwendung für das betriebliche Ausbildungspersonal in KKUs

Franziska Günther <sup>1</sup>, Jörg Neumann<sup>2</sup>, Robert Lorenz <sup>3</sup> und Annegret Umlauf<sup>4</sup>

**Abstract:** Klein- und Kleinstunternehmen (KKUs) stehen im Zuge von Aus- und Weiterbildung vor dem Problem einer stark limitierten Personalstruktur, die einen teils mehrtägigen Arbeitsausfall, z.B. aufgrund von Weiterbildung, nicht gestattet. Zudem ist zu beobachten, dass gerade bei KKUs, die vom Einsatz von digitalen Medien in der Ausbildung profitieren würden, die Einbindung derer nur zögerlich vorangeht. Eine orts- und zeitflexible Möglichkeit zur Weiterbildung stellt das Lernen durch eine App dar. Im Folgenden wird daher die Entwicklung der App „Ausbilderakademie.digital“ vorgestellt. Durch die App soll das betriebliche Ausbildungspersonal in KKUs lernen, wie Medien in der Ausbildung eingesetzt werden können. Im Zuge dessen wird daher auch die Medienkompetenz der Zielgruppe trainiert. Im Beitrag wird auf die didaktischen Aspekte, vor allem vor dem Hintergrund von Mobile Learning und Microlearning, eingegangen.


**Keywords:** Mobile Learning, Microlearning, Medienkompetenz, Ausbildungspersonal, berufliche Bildung.

## 1 Einleitung und Motivation


Dem Einsatz digitaler Medien in der Ausbildung stehen laut einer Studie des Bundesinstituts für Berufsbildung (BIBB) die meisten der befragten AusbilderInnen offen gegenüber [Hä18]. Knapp 70% der Befragten stimmen der Aussage zu, dass durch den Einsatz digitaler Medien die Qualität der Ausbildung gesteigert werden kann. Circa 80% sind sich darüber einig, dass sich vor allem die Attraktivität der Ausbildung erhöht. Jedoch fällt auf, dass bei der Einbindung digitaler Medien in den Alltag der beruflichen Ausbildung auf Seiten des Ausbildungspersonals Hemmnisse bestehen, diese einzusetzen. So sind sich über die Hälfte der befragten AusbilderInnen unsicher bzw. unentschieden darüber, was der Einsatz digitaler Medien in der Ausbildung bringt. Über 70% sind mindestens unentschieden bei der Aussage, ob die ohnehin hohe zeitliche Belastung als

---

<sup>1</sup> TU Dresden, Medienzentrum, Strehleener Straße 22/24, 01069 Dresden, franziska.guenther1@tu-dresden.de,

 <https://orcid.org/0000-0002-9022-6622>

<sup>2</sup> TU Dresden, Medienzentrum, Strehleener Straße 22/24, 01069 Dresden, joerg.neumann@tu-dresden.de

<sup>3</sup> TU Dresden, Medienzentrum, Strehleener Straße 22/24, 01069 Dresden, robert.lorenz4@tu-dresden.de,   
<https://orcid.org/0000-0002-8776-1798>

<sup>4</sup> Handwerkskammer Dresden, Bildungszentrum, Am Lagerplatz 8, 01099 Dresden, annegret.umlauft@hwk-dresden.de

AusbilderIn den Einsatz zusätzlich erschwert. Knapp 80% der Befragten sind sich darüber einig, dass fehlendes Know-How von AusbilderInnen den Einsatz digitaler Medien in der Ausbildung verhindert.

Weiterbildungen im Bereich der Medienkompetenz und des Medieneinsatzes können dem Ausbildungspersonal helfen, diese Wissenslücken zu schließen. Allerdings ist die Personalaufstellung gerade in Klein- und Kleinstbetrieben des Handwerks so gering, dass ein mehrtätiger Ausfall im Rahmen einer Weiterbildung nur für die wenigsten Unternehmen in Frage kommt. Orts- und zeitflexible Angebote für das Ausbildungspersonal stellen eine Möglichkeit dar, Wissen bedarfsgerecht zu vermitteln. Daher ist das Lernen mit dem Smartphone, z.B. über Apps, gerade für AusbilderInnen von hohem Interesse.

Wie muss eine solches Lernangebot hinsichtlich praktischer und didaktischer Anforderungen aussehen, um von der Zielgruppe angenommen zu werden? Im folgenden Beitrag soll diese Frage anhand der Entwicklung der App „Ausbilderakademie.digital“ beispielhaft beantwortet werden. Die App wird zusammen mit der Handwerkskammer Dresden am Medienzentrum der TU Dresden entwickelt. Durch die Nutzung der App soll das Ausbildungspersonal auf niedrigschwelligem Weg für die Themen Medienkompetenz und Medieneinsatz in der Ausbildung sensibilisiert werden. Im Folgenden werden verschiedene grundlegende didaktische Entscheidungen vor dem Hintergrund mobilen Lernens in der betrieblichen Bildung diskutiert.

## **2 Mobile Learning in der betrieblichen Bildung**

Mobile Learning ist nach Bartelsen (2011) das „Lernen mit mobiler Computer- und Telekommunikationstechnologie, sofern dabei Lehr-Lern-Szenarien zum Einsatz kommen, die speziell auf mobile Endgeräte angepasst sind.“ [Ba11, S. 4]. Vor allem in Lernszenarien ohne Bezug zum Lernort oder Nutzungskontext ist Mobile Learning ein mittlerweile in der Erwachsenenbildung „weit verbreitetes Nutzungsszenario“ [Br17, S. 98]. Also besonders da, wo Lerninhalten auf mobile Endgeräte übertragen werden, um einen orts- und zeitunabhängigen Zugriff zu ermöglichen [Br17].

Laut dem IW-Personalpanel 2018 nutzen allerdings lediglich 34,4% digitale Lernangebote an mobilen Endgeräten zu Weiterbildungszwecken in Unternehmen [F119]. Die am meisten genutzte Form digitaler Lernangebote sind digitale Texte, z.B. als PDF bereitgestellt. Dazu konträr stehen die Zahlen der Trendstudie mmb Learning Delphi 2019/2020. Diese besagen, dass vor allem mobile Anwendungen, als Lernform für das betriebliche Lernen in Unternehmen, in den nächsten drei Jahren kommerziell sehr erfolgreich sein werden [Mm20]. Zwischen Erwartung und Verbreitung besteht demnach noch eine Kluft. Vor allem die hohe Selbstlernkompetenz, die für die Nutzung von mobilen Lernangeboten notwendig ist, ist nach einem Literaturreview von Seufert und Meier ein Grund, warum Mobile Learning in Unternehmen noch nicht verbreitet ist

[SM18]. Motivierende Elemente, um die Selbstlernkompetenz zu steigern, können kurze und prägnante Lerninhalte sowie spielerische Elemente wie Quizze sein. Weiterhin haben Seufert und Meier nach der Sichtung verschiedener Studien zum mobilen Lernen festgestellt, dass an Hochschulen die didaktische Einbindung mobiler Lernangebote mehr Bedeutung findet als in Unternehmen. Hier steht zumeist die technische Umsetzung im Vordergrund [SM18].

Didaktische Überlegungen und die Einbindung mobiler Lernangebote in geeignete Lernszenarien sind aber notwendig, um den Mehrwert des Angebots für die Lernenden sichtbar zu machen. Zudem hat sich gezeigt, dass sich eine Hinführung der Lernenden zum Lernangebot positiv auf die zukünftige Nutzung auswirkt [BT18]. Dies kann beispielweise über ein Blended Learning Szenario geschaffen werden. Das didaktische Design, welches im Folgenden vorgestellt wird, wurde auf Grundlage der Erkenntnisse von mobilem Lernen in der betrieblichen Bildung entwickelt.

### 3 Didaktisches Design

Um das didaktische Design der App zu entwickeln, wurde der Konzeptionsrahmen von Beutner und Teine (2018) für die Gestaltung zielgruppengerechter Lernangebote im mobilen Bereich genutzt [BT18]. Des Weiteren wird gezeigt, wie durch Microlearning mit Lernkarten der Konzeptionsrahmen umgesetzt werden kann.

#### 3.1 Konzeptionsrahmen

**Auswahl der Lerninhalte:** Grundlage für die Auswahl der Lerninhalte ist ein Kompetenzrahmen, der sich aus verschiedenen etablierten Modellen zur Förderung der Medienkompetenz und des Medieneinsatzes zusammensetzt. Auf dieser Grundlage wiederum werden bestehende Lerninhalte, vorrangig OER, recherchiert und eingebunden, wie z.B. Materialien aus dem Vorgängerprojekt medienBAR [Um18].

**Zielbestimmung:** Durch die Nutzung der App soll das betriebliche Ausbildungspersonal für die Themen Medienkompetenz und Medieneinsatz in der Ausbildung sensibilisiert werden. Nach Seufert und Meier hat insbesondere „mobiles Lernen Potenzial für die Entwicklung von digitalen Kompetenzen (bzw. Kompetenzen für eine digitale Lebens- und Arbeitswelt)“ [SM18, S. 892].

**Niveaubestimmung:** Da das Ausbildungspersonal für die genannten Themen niedrigschwellig sensibilisiert werden soll, werden die Anforderungen geringgehalten. Die einzelnen Lerneinheiten können für sich stehend absolviert werden, es bestehen daher keine komplexen Zusammenhänge zu anderen Lerneinheiten. Eine Steigerung des Anforderungsniveaus ist im aktuellen Szenario nicht vorgesehen.

**Einbindung von Vermittlungs-/Sinneskanälen:** Medienformate wie Texte, Videos, Podcasts oder Quizze werden in die Lerneinheiten eingebunden.

**Mediale und methodische Repräsentation:** Die in sich abgeschlossenen Lerneinheiten sind jeweils in Themenbereiche gegliedert, die wiederum zu übergeordneten Kompetenzbereichen gehören. Die Kompetenzbereiche wurden auf Grundlage verschiedener Kompetenzrahmen ([Ba96] [KM00] [B100]) entwickelt.

**Authentizitätsgrad:** In jeder Lerneinheit sind Best Practices oder Anwendungsszenarien eingebettet, um den Lernenden den Inhalt so praxisrelevant wie möglich nahezubringen. Diese werden mit Hilfe der Erprobungsgruppen zusammen entwickelt.

**Kontroll- und Testmöglichkeiten:** Jede Lerneinheit besteht aus einem Quiz (z.B. Multiple Choice oder Zuordnungen) und einer Zusammenfassung.

### 3.2 Microlearning mit Lernkarten

Mit Microlearning wird versucht, Lerninhalte und -aktivitäten so abzubilden, dass die Lerndauer kurz gehalten werden kann. Der Ansatz wird oft mit dem Learning Nugget Ansatz von Bailey (2006) gleichgesetzt. Die Lerneinheiten sind jeweils abgeschlossen und bestehen zumeist aus einem Lernziel, einem praxisnahen Beispiel sowie aus Aufgaben und verschiedenen Medienformaten [Ba06]. Für Hug (2018) ist Microlearning auch „das Lernen mit [...], kleinen Lerneinheiten, mobilen Geräten oder Mikroinhalten.“ [Hu18, 2018, S. 322].

Damit steht Microlearning auch im Zusammenhang mit Mobile Learning. Weiterhin stellt Hug (2018) fest, dass beide Konzepte ähnliche Didaktiken teilen und sich auch einander, z.B. in technischen Aspekten durch kleine Bildschirme, bedingen. Darüber hinaus lässt sich feststellen, dass „die Aufmerksamkeitsspannen und Zeitperioden der meisten mobilen Anwendungen relativ kurz“ sind [Hu18] und Microlearning demnach den Gewohnheiten vieler App-Nutzenden entspricht.

Mikrolerneinheiten in mobilen Anwendungen können anschaulich durch Lernkarten dargestellt werden (Abbildung 1). Der Inhalt beschränkt sich dabei auf ein bis zwei Medienformate pro Karte, wie z.B. einen kurzen Text, ein Quiz, ein Video oder ein Bild. Um zur nächsten Lernkarte zu gelangen wird die Swipe-Geste am Smartphone genutzt.

## 4 Entwicklung der Lerneinheiten

Die Umsetzung des Konzeptionsrahmens und der Mikrolerneinheiten erfolgt zunächst auf Grundlage des Phasenschemas nach Beutner und Teine [BT18]. Daraufhin folgt die Erstellung von Mock-ups (Abbildung 1), die der Erprobung und als Grundlage für die technische Umsetzung dienen. Ähnlich dem Micro-Unit Phasenschema nach Beutner und

Teine [BT18] besteht eine Lerneinheit aus Lernzielen, einem praxisnahen Beispiel, dem Lerninhalt, einem Quiz und einer Zusammenfassung. Die Länge einer Lerneinheit sollte im Durchschnitt nicht mehr als fünf Minuten betragen.

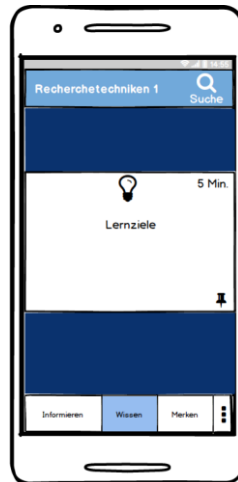


Abb. 1: Mock-up der Lerneinheit mit der Lernkarte „Lernziel“

## 5 Ausblick

Die Wirksamkeit der Kombination von Mobile Learning und Microlearning zur Förderung der Medienkompetenz und des Medieneinsatzes durch das Ausbildungspersonal in der Ausbildung wird in den nächsten Phasen des Vorhabens erprobt und evaluiert werden. Erprobungsgruppen werden durch die Bedienung von interaktiven Mock-ups den App-Entwurf bezüglich Bedienungsfreundlichkeit und Inhalt in mehreren Iterationszyklen bewerten. Die nötigen Daten werden mittels Methoden wie Lautes Denken und Befragungen mit Fragebögen erhoben und qualitativ wie quantitativ ausgewertet. Nach den Auswertungen folgt jeweils eine Anpassung der App an die Anforderungen der Zielgruppe. Parallel dazu wird der Kompetenzrahmen weiter ausgebaut, um auf dessen Grundlage das Feinkonzept der Lerneinheiten zu erstellen. Weiterhin wird ein Blended Learning Szenario entwickelt, um die Zielgruppe in den Präsenzphasen an die Nutzung der App heranzuführen. Der hier dargestellte Ablauf zur Entwicklung einer App soll als Best Practice Beispiel für weitere Projekte im Bereich der beruflichen Bildung dienen.

## Literaturverzeichnis

- [Ba96] Baacke, D.: Medienkompetenz – Begrifflichkeit und sozialer Wandel. In (von Rein, A. Hrsg.): Medienkompetenz als Schlüsselbegriff. Deutsches Institut für

- Erwachsenenbildung (DiE), S. 112–124, 1996, [https://www.die-bonn.de/esprid/dokumente/doc-1996/rein96\\_01.pdf#page=111](https://www.die-bonn.de/esprid/dokumente/doc-1996/rein96_01.pdf#page=111), Stand: 12.08.2020.
- [Ba06] Bailey, C.; Zalfan, M. T.; Davis, H. C.; Fill, K.; Conole, G.: Panning for gold: Designing pedagogically-inspired learning nuggets. *Educational Technology and Society*, 9(1), S. 113–122, 2006.
- [Ba11] Bartelsen, J.: Mobile learning – ein Überblick. Arbeitspapiere der Nordakademie Nr. 2011/06, 2011, <https://www.econstor.eu/bitstream/10419/67093/1/685361942.pdf>, Stand: 12.08.2020.
- [BT18] Beutner, M.; Teine, M.: Mobile Learning für alle. In (de Witt, C; Gloerfeld, C. Hrsg.): *Handbuch Mobile Learning*. Springer Fachmedien, S. 385–407, 2018, [https://doi.org/10.1007/978-3-658-19123-8\\_20](https://doi.org/10.1007/978-3-658-19123-8_20).
- [Bl00] Blömeke, S.: *Medienpädagogische Kompetenz: Theoretische und empirische Fundierung eines zentralen Elements der Lehrerbildung*. KoPäd-Verlag, 2000.
- [Br17] Bremer, C.: Mobiles Lernen in der Erwachsenenbildung. Optionen für Lernszenarien. In (Thissen, F. Hrsg.): *Lernen in virtuellen Räumen*. De Gruyter, S. 88–109, 2017, <https://doi.org/10.1515/9783110501131-007>.
- [Fl19] Flake, R.; Malin, L.; Meinhard, D. B.; Müller, V.: *Digitale Bildung in Unternehmen – Wie KMU E-Learning nutzen und welche Unterstützung sie brauchen*. Institut der deutschen Wirtschaft Köln e. V., 2019. [https://www.iwkoeln.de/fileadmin/user\\_upload/Studien/Gutachten/PDF/2019/KOFA-Studie\\_3\\_2019\\_Digitale\\_Bildung.pdf](https://www.iwkoeln.de/fileadmin/user_upload/Studien/Gutachten/PDF/2019/KOFA-Studie_3_2019_Digitale_Bildung.pdf), Stand: 12.08.2020.
- [KM16] Kultusministerkonferenz: *Kompetenzen in der digitalen Welt*. Kultusministerkonferenz. 2016, [https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/2016\\_12\\_08-KMK-Kompetenzen-in-der-digitalen-Welt.pdf](https://www.kmk.org/fileadmin/Dateien/pdf/PresseUndAktuelles/2016/2016_12_08-KMK-Kompetenzen-in-der-digitalen-Welt.pdf), Stand: 12.08.2020.
- [Hä18] Härtel, M.; Brüggemann, M.; Sander, M.; Breiter, A.; Howe, F.; Kupfer, F.: *Digitale Medien in der betrieblichen Berufsbildung: Medienaneignung und Mediennutzung in der Alltagspraxis von betrieblichem Ausbildungspersonal*. Bundesinstitut für Berufsbildung, 2018.
- [Hu18] Hug, T.: Mikrolernen und mobiles Lernen. In C. de Witt & C. Gloerfeld (Hrsg.), *Handbuch Mobile Learning*. Springer Fachmedien Wiesbaden, S. 321–340, 2018, [https://doi.org/10.1007/978-3-658-19123-8\\_17](https://doi.org/10.1007/978-3-658-19123-8_17).
- [Mm20] mmb Institut: *Mmb Trendmonitor 2019-2020*. mmb Institut – Gesellschaft für Medien- und Kompetenzforschung mbH, 2020, [https://www.mmb-institut.de/wp-content/uploads/mmb-Trendmonitor\\_2019-2020.pdf](https://www.mmb-institut.de/wp-content/uploads/mmb-Trendmonitor_2019-2020.pdf), Stand: 12.08.2020.
- [SM18] Seufert, S.; Meier, C.: Empowerment-orientierte Kompetenzentwicklung auf der Grundlage von Mobile Assisted Seamless Learning. In (de Witt, C.; Gloerfeld, C. Hrsg.): *Handbuch Mobile Learning*. Springer Fachmedien, S. 879–899, 2018, [https://doi.org/10.1007/978-3-658-19123-8\\_41](https://doi.org/10.1007/978-3-658-19123-8_41).
- [Um18] Umlauf, A.; Zillger, J.; Neumann, J.: *medienBAR – medienpädagogische Basisqualifizierung für das Ausbildungspersonal im Handwerk*. Schlussbericht, Handwerkskammer Dresden, 2018, <https://doi.org/10.2314/GBV:1042348553>.



## Objektorientierte Programmierung - Kompetenzerwerb im Mastery Model des Inverted Classroom

### Aufbau und Akzeptanz der Kursumsetzung

Annett Thüring<sup>1</sup> und Kathrin Jäger<sup>2</sup>

**Abstract:** Das Modul Objektorientierte Programmierung wird an der Martin-Luther-Universität Halle-Wittenberg für ca. 300 Studierende im ersten Semester klassisch als Präsenzvorlesung mit Übung angeboten. Vor dem Hintergrund steigender und hoher Abbruchquoten wurde für diese Lehrveranstaltung ein lernerzentriertes Blended-Learning-Konzept entwickelt, das auf dem Inverted Classroom Mastery Modell (ICMM) basiert. Das didaktische Konzept setzt an der Praxis an, fokussiert auf den Kompetenzerwerb und orientiert auf eine Förderung selbstregulierten Lernens unter Einbeziehung der Rolle des Lehrenden in diesem Ansatz. Nach einer ersten Testphase im WS 2019/20 können Ergebnisse des Einsatzes des ICMMs und praktische Erfahrungen bei der Nutzung der Lernumgebung positiv diskutiert werden.

**Keywords:** Programmierkurs, Inverted Classroom Mastery Modell, Kompetenzmessung

## 1 Hintergrund

Objektorientierte Programmierung (OOP) ist eine Lehrveranstaltung für Studierende des ersten Semesters, die an der Martin-Luther-Universität Halle-Wittenberg klassisch in Präsenz sowohl Vorlesung als auch Übung umfasst. Studierende unterschiedlicher Studienrichtungen wie z.B. Informatik, Mathematik, Wirtschaftsinformatik, Lehramt Informatik und Physik können das Modul obligatorisch oder wahlobligatorisch in ihr Studium einbringen. In den letzten Jahren wurden sinkende Lernerfolge und in der Fortführung hohe Kursabbrüche seitens der Studierenden verzeichnet. Diese werden vor allem auf die zunehmende Heterogenität der Zielgruppe bzgl. des individuellen Vorkenntnisstandes aber auch auf die Bedürfnisse und Wünsche der Studierenden nach individualisierten Studienwegen und flexiblen Studienmodellen zurückgeführt. Eine vorangestellte Analyse des Modells Vorlesung und Übung in Präsenz ergab, dass eine passive Vorlesung zur Vermittlung von Fähigkeiten zudem ungeeignet ist. Vor diesem Hintergrund fokussiert eine Umstrukturierung des Moduls OOP die Einbeziehung und Passung zeitgemäßer Lehr-/Lernformen.

---

<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, Von-Seckendorff-Platz 1, 06120 Halle (Saale), annett.thuering@informatik.uni-halle.de

<sup>2</sup> Martin-Luther-Universität Halle-Wittenberg, Zentrum für multimediales Lehren und Lernen (LLZ), Hoher Weg 8, 06120 Halle (Saale), kathrin.jaeger@llz.uni-halle.de

## 2 Ziele

Hinsichtlich des individuellen Vorkenntnisstandes der Studierenden können angebotene Lehrveranstaltungen nur effizient durchgeführt werden, wenn ein möglichst homogenes Wissensniveau aller Lernenden erreicht werden kann. In der Praxis haben sich hierzu Konzepte mit einer Integration von E-Learning-Anteilen in Form von Blended-Learning bewährt. Zur Individualisierung des Modulangebotes basiert die vorgestellte Umstrukturierung der Lehrveranstaltung auf einem lernerzentrierten dreiphasigen Inverted Classroom Mastery Modell (ICMM) [Ha15], das neben onlinebasierter Wissensvermittlung und problembasierten Austausch- und Übungsphasen in Präsenz, formative Assessments integriert und Lernerpräferenzen berücksichtigt. Unterstützend werden universelle Ansätze aus Motivationsmodellen<sup>3</sup> herangezogen sowie Planungs- und Regulationshilfen für selbstreguliertes Lernen<sup>4</sup> implementiert. Umsetzungen stützen sich auf didaktische Konzepte mit einer Balance zwischen Instruktion und Konstruktion sowie auf eine Passung der Lehr-/Lernformen und Interventionen zu einer stufenweisen Vermittlung der Fachkompetenz. Dabei werden die technologiegestützten Elemente auf das Verstehen immer komplexer werdender Problemstellungen ausgerichtet. Ein ähnlicher Ansatz wird auch von Fässler verfolgt [Fä07], wobei hier Studierende nur am Ende Einblick in den Kompetenzstand erhalten. Ziel dieses Kurses ist es, den Lernerfolg schrittweise zu ermöglichen, zu messen und zudem transparent den Studierenden den Lernstand sichtbar zu machen. Rückmeldungen zum Lernstand fokussieren die Selbstregulation, sodass das Kursziel, die Fähigkeit des Programmierens, erreicht werden kann.

## 3 Das didaktische Konzept

Das didaktische Konzept orientiert auf eine verstärkte Einbeziehung onlinebasierter digitaler Elemente, wobei ein intensiver Analyseprozess für die Auswahl der Inhalte und Aktivitäten vorangeht. Die Grundstruktur der 15 Kurseinheiten im Semesterzyklus wird beibehalten und zur Planung die umfassende didaktische Designplanung nach Reinmann [Re15] herangezogen. Im ersten Schritt erfolgte eine Formulierung wissens- und kompetenzorientierter Lernziele<sup>5</sup> sowohl der gesamten Lehrveranstaltung als auch jeder Kurseinheit. Die Zuordnung der Lern- und Kompetenzziele basiert auf der Bloom'schen Lernzieltaxonomie [AK01] sowie den von Weicker et al. [WWD06] formulierten Schlüsselkompetenzen für Informatiker.

---

<sup>3</sup> Grundlegende Ansätze im Modulablauf integrieren Aspekte aus dem ARSCS-Modell von Keller [KK87], dem Motivationsprogramm von Fries [Fr02] und dem ECOLE-Modell (Emotionale und kognitive Aspekte des Lernens) von Gläser und Zikuda [Gl05].

<sup>4</sup> Studierende benötigen bei selbständiger Wissensaneignung in Online-Phasen Selbstregulationsstrategien. [ZS01]

<sup>5</sup> Empfehlungen der Gesellschaft für Informatik für Bachelor- und Masterprogramme im Studienfach Informatik in denen Lernziele, Kompetenzen und Empfehlungen im Studienfach Informatik verankert sind. [Zu16]

Die Lern- und Kompetenzziele geben dem Lehrenden eine Grundlage zur Strukturierung, Selektion und Reduktion der Inhalte. Alle Inhaltselemente und Aufgaben des Kurses werden danach ausgewählt und zugeordnet und die Prüfungsaufgaben abgeleitet<sup>6</sup>. Sichtbare Lern- und Kompetenzziele im Modulablauf, sowie eine klare Struktur bieten dem Studierenden zudem Transparenz, Verbindlichkeit und Orientierungshilfe.

Für die eigentliche Inhaltserstellung und den Ablauf der verschiedenen Lernphasen werden unterschiedliche Modelle und Methoden herangezogen, um Aufmerksamkeit und Selektion sowie Verarbeitung und Integration des Wissens in die individuelle Wissensstruktur zu ermöglichen und zu fördern. Als Hauptlehrmethode über den gesamten Kurs wird das Cognitive Apprenticeship-Modell (CAM) [CBN89] genutzt. Dieses Modell wurde an die traditionelle Handwerkslehre angelehnt und als schulisches Unterrichtsmodell entwickelt. Es geht davon aus, dass der Lehrende für ein Problem oder eine Aufgabe die Lösung vorführt und gleichzeitig die kognitiven aber auch metakognitiven Schritte artikuliert. Lernende versuchen im Anschluss selbst Probleme gleicher Klasse auf die gleiche Art zu lösen und der Lehrende zieht sich zurück. Im Vergleich weist das CAM eine sehr viel höhere und auch konkretere Intensität an instruktionaler Unterstützung auf als andere Methoden wie z. B. Anchored Instruction. Dieses Ausmaß an Unterstützung steht im universitären Lehrkontext zumeist nicht im Fokus. Für Studierende im Erstsemester, die noch wenig Erfahrung mit selbstreguliertem Lernen haben, kann diese Methode gewinnbringend sein. Eine Diskussion über die Handlungsschritte zwischen Lehrenden und Studierenden aber auch zwischen Studierenden untereinander ermöglicht zudem in die jeweilige Fachkultur hineinzuwachsen und zukünftiges funktionales Verhalten bezüglich der Programmieraufgaben zu erlernen. Abb. 1 zeigt die instruktionale Sequenz der didaktischen Bausteine *modeling*, *scaffolding*, *fading*, *coaching*, *exploration*, *articulation* und *reflection* über den Kursverlauf (Online und in Präsenz).

### Umsetzung im Lernmanagementsystem ILIAS<sup>7</sup>

Alle Online-Kursmaterialien<sup>8</sup> werden im ILIAS zur Verfügung gestellt. Die Kurseinheiten sind identisch aufgebaut (Abb. 1) und werden tutoriell und technisch begleitet. Formative Tests sind direkt in den Kurseinheiten implementiert und setzen am Mastery Learning an. Die Studierenden erhalten sofort Rückmeldung zu ihrem Lernstand. Zusätzlich werden sie durch fachlich betreute Foren, ein ausgeprägtes Feedback und ein Rückkopplungssystem unterstützt. Zum Lösen der Programmieraufgabe steht eine webbasierte Arbeitsumgebung (YAPEX)<sup>9</sup> zur Verfügung. Abgeschlossen wird jede Kurseinheit mit einem Wissenstest, fortschreitend mit einem Kompetenztest. Nach jeder Kurseinheit ist eine Umfrage implementiert, die der Evaluation dient und zudem Aspekte zur Metakognition enthält, um bei den Studierenden eine aktive Überwachung, Bewertung und

<sup>6</sup> Passung von Lernzielen, Inhalten und Prüfungsaufgaben (constructive Alignment) [Bi96]

<sup>7</sup> ILIAS - Integriertes Lern-, Informations- und Arbeitskooperations-System ist ein Open-Source Learning Management System. [<https://www.ilias.de/>]

<sup>8</sup> Demo-Kurs abrufbar unter: <https://bit.ly/2UJfSud>. (Nutzer: ICMMDemo, Passwort: ICMM\_ilias2020)

<sup>9</sup> YAPEX – (Yet Another Practical EXercise Platform) ist eine von der Uni Halle am Institut für Informatik entwickelte webbasierte Arbeitsumgebung zum Bearbeiten von Programmieraufgaben. [Dä16]

Regulation des eigenen Lernprozesses anzuregen. Eine elektronische Prüfung lässt abschließend zu, aus Lernerfolgen Rückschlüsse auf erworbene Kompetenzen zu ziehen.

Komponenten / inhaltliche Elemente Online/Präsenz	Umsetzung Elemente Cognitive Apprenticeship	Lernziele Taxonomie
<b>Online LMS ILIAS (1.-15. Kurseinheit)</b> didaktisch und unterschiedlich medial aufbereitete Inhalte (Videoinput, Text, Bild), Codebeispiele, eingebettete Übungen etc. mit Kohärenzhilfen zum selbstregulierten Lernen, Integration von Feedback- und Kommunikationstools	<b>Content</b> Vermittlung von Fakten, Konzepten und prozeduralem Wissen <b>Situated Learning</b> Lernende arbeiten mit realistischen und authentischen Problemen	Wissen aneignen verstehen abrufen
<b>Präsenz (1.-15. Kurseinheit)</b> kleinere Übungsaufgaben zur Förderung verstehenden Lernens, Möglichkeit zum intensiven Diskurs mit Mitstudierenden, dem Lehrenden und erfahrenen Studierenden	<b>Modeling and Explaining</b> Lehrender demonstriert eine Problemlösung und kommentiert die Strategie <b>Scaffolding</b> Lehrender stellt eine Orientierungsgrundlage bereit, Studierende lösen ein ähnliches Problem	Wissen aneignen verstehen abrufen anwenden
<b>Online Programmieraufgabe</b> LMS ILIAS u. YAPEX (1.-11. Kurseinheit) komplexe Anwendungsaufgabe zum aktuellen Kompetenzstand, Bearbeitung in einer webbasierten Entwicklungsumgebung (YAPEX) mit sofortiger Rückmeldung	<b>Fading</b> Lehrender zieht sich bei steigender Kompetenz der Lernenden zurück <b>Coaching</b> Problemlösung der Lernenden, Lernende erhalten Hinweise und Feedback vom Lehrenden	Wissen verstehen abrufen anwenden analysieren
<b>Online Praxisprojekt</b> LMS ILIAS u. YAPEX (12.-15. Kurseinheit) Entwickeln einer Software im realen Kontext (Programmieren kleiner lauffähiger Programme und Vorstellen des Gesamtprojektes)	<b>Exploration</b> Problemlösungskompetenz wird auf andere Bereiche übertragen bzw. Variationen vom Lernenden ausprobiert <b>Articulation</b> Lernende müssen ihre Strategien kommentieren und Entscheidungen begründen	Wissen abrufen anwenden analysieren synthetisieren reflektieren
<b>Online Formative Tests</b> LMS ILIAS u. YAPEX (1.-15. Kurseinheit) Selbsttest zur Lernstandsüberprüfung mit sofortigem Feedback <b>Online Evaluation</b> LMS ILIAS (1.-15. Kurseinheit) Umfrage (geschlossene und offene Fragen)	<b>Reflection</b> Lernende haben Einsicht in ihren Entwicklungsfortschritt und können ihn analysieren und vergleichen Lernende reflektieren den Lernprozess und treffen ggf. Maßnahmen zur Selbstregulation	Wissen abrufen anwenden analysieren reflektieren

Abb. 1: Umsetzung des Moduls OOP im ICMM – Elemente CAM

## 4 Ergebnisse der Testphase des ICMM

Die Veranstaltung im ICMM wurde erstmalig im Wintersemester 2019/20 durchgeführt und evaluiert. Die Studierenden konnten selbst wählen, ob sie an der klassischen Veranstaltung teilnehmen (217 Studierende) oder das ICMM-Format (20 Studierende) wählen. Im Semesterverlauf wurden über Beobachtungen, anonyme Online-Umfragen je Kurseinheit und eine abschließende Evaluation unterschiedliche lernrelevante Aspekte untersucht. Nachfolgende Betrachtungen richten sich vor allem auf die Motivation der Studierenden und den Lernerfolg (Zuwachs an Kompetenzen) über den Kursverlauf. Kein Studierender hat zur klassischen Lehrveranstaltung gewechselt. Zwei Studierende (OOP als wahlobligatorisches Modul) brachen den Kurs nach Erreichen des ersten Kompetenzlevels ab. Alle Online-Kurseinheiten wurden intensiv bearbeitet (gemessen über Lernfortschrittskontrolle im ILIAS, anonymisierte Zeit- und Seitenerfassung, sowie über die Ergebnisse aus formativen Assessments). Gruppenarbeiten zeigten eine hohe Aktivität in Form der Beteiligung aller Studierenden am Endergebnis. Die Mehrzahl der Studierenden war an einer partizipativen Mitgestaltung und Weiterentwicklung der Lernumge-

bung interessiert. Dies äußerte sich darin, dass die Studierenden freiwillig und eigenständig ergänzende Testfälle für YAPEX-Aufgaben erstellten und die Kursgestaltung durch Erweiterung der Online-Kurseinheiten mit praxisnahen Beispielen bereicherten. Zur Bewertung des Kompetenzzuwachses wurden die Ergebnisse der Online-Tests und Online-Programmieraufgaben herangezogen. Eine quantitative Analyse ergab, dass 39% der Studierenden den klassischen Kurs abgebrochen haben, aber nur 10% der Studierenden den Kurs im ICMM. Die elektronische Abschlussprüfung misst inwieweit die Studierenden objektorientierte Systeme modellieren und Algorithmen in eine Programmiersprache umsetzen, sowie Testfälle formulieren können. Abb. 2 zeigt im Vergleich den Lernerfolg für den Kurs im ICMM und den klassischen Kurs für Klausur 1. Es meldeten sich anteilig mehr Studierende aus dem Kurs im ICMM zur ersten Klausur an. Alle Studierenden haben die Klausur bestanden, die schlechteste Note war 3,3 (ein Studierender). Aus dem klassischen OOP-Kurs haben dagegen 10,68% der Studierenden die erste Klausur nicht bestanden, sieben Studierende erhielten zudem die Note 3,7 und schlechter.

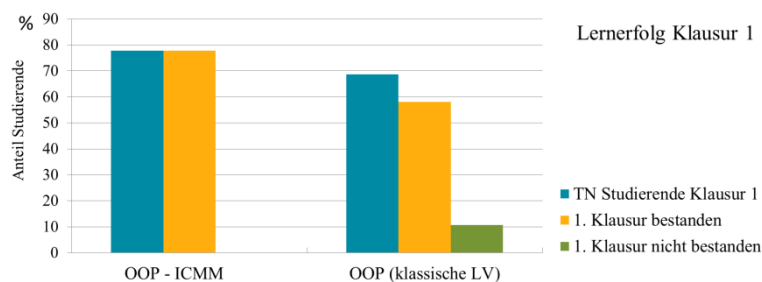


Abb. 2: Klausurteilnahme und summativer Lernerfolg (Klausur 1)

## 5 Fazit

Die Ergebnisse des Einsatzes des ICMMs und praktische Erfahrungen bei der Nutzung der Lernumgebung OOP zeigen, dass Studierende die Kursinhalte mit hoher Motivation ergebnisorientiert bearbeiteten. Positiv hervorzuheben sind die geringe Abbruchquote und der höhere Lernerfolg im Vergleich zur klassischen Veranstaltung. Der Lernerfolg, formativ und summativ, lässt zudem die Schlussfolgerung zu, dass die Cognitive Apprenticeship-Methode für den Erwerb der Kompetenz der Programmierfähigkeit einen Mehrwert zu bieten scheint. Die didaktische und mediale Gestaltung der Lernumgebung des ICMM wird zusätzlich durch die Einstellungsakzeptanz der Studierenden unterstützt. Lehrende müssen jedoch bei der Umsetzung des CAMs eine sehr viel höhere und auch konkretere Intensität an instruktionaler Unterstützung anbieten. Nachhaltig ist die Wiederverwendung und unkomplizierte Modifizierung sowie Ergänzung des Materials. Aufgrund der positiven Bilanz der Testphase ist vorgesehen, die Veranstaltung weiter auszubauen und für alle Studierenden anzubieten. Hierbei sind eine ATI-Studie und eine statistische Auswertung basierend auf dem unterschiedlichen Vorwissensniveau der Studierenden hinsichtlich der Wirksamkeit gegen Abbruch und zum Lernerfolg geplant.

## Literaturverzeichnis

- [AK01] Anderson, L. W., Krathwohl, D. R.: A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, Boston: Pearson Education Group, 2001.
- [Bi96] Biggs, J.: Enhancing teaching through constructive alignment, in: Higher Education 32, 347–364, 1996.
- [CBN89] Collins, A., Brown, J. S., Newman, S. E.: Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L.B. Resnick (Hrsg.), Knowing, learning and instruction: Essays in honor of Robert Glaser. Hillsdale, NJ: Erlbaum, S. 453-494, 1989.
- [Dä16] Dähne, J.: YAPEX, 2016. <https://yapex.informatik.uni-halle.de/>, Stand: 28.03.2020.
- [Fä07] Fässler, L. E.: Das 4-Schritte-Modell – Grundlage für ein kompetenzorientiertes E-Learning (Dissertation Nr. 17521), ETH Zürich, 2007.
- [Fr02] Fries, S.: Wollen und Können: Ein Training zur gleichzeitigen Förderung des Leistungsmotivs und des induktiven Denkens, Münster, Waxmann 2002.
- [Gl05] Gläser-Zikuda, M. et.al.: Promoting students' emotions and achievement – Instructional design and evaluation of the ECOLE-approach. Learning and Instruction, Volume 15, Issue 5, S. 481-495, 2005.
- [Ha15] Handke, J.: Digitalisierung der Hochschullehre - Welche Rolle spielt das Inverted Classroom Model dabei?, In Freisleben-Teutscher C. F. / Gruber W. / Haag J. / Weißböck J. (Hrsg.) Neue Technologien – Kollaboration – Personalisierung, Tagungsband zum 3. Tag der Lehre, 2015.
- [KK87] Keller, J. M., Kopp, T. W.: An application of the ARCS model of motivational design. In Reigeluth, C. M. (Hg.): Instructional theories in action. Lessons illustrating selected theories and models, Hillsdale, NJ: Erlbaum, S. 289–320, 1987.
- [Re15] Reinmann, G.: Studententext Didaktisches Design. Hamburg, 2015. Abruf unter: [https://gabi-reinmann.de/wp-content/uploads/2013/05/Studententext\\_DD\\_Sept2015.pdf](https://gabi-reinmann.de/wp-content/uploads/2013/05/Studententext_DD_Sept2015.pdf), Stand: 28.03.2020.
- [WWD06] Weicker, N., Draskoczy, B., Weicker, K.: Fachintegrierte Vermittlung von Schlüsselkompetenzen der Informatik. In: Forbrig, P., Siegel, G. & Schneider, M. (Hrsg.), HDI 2006: Hochschuldidaktik der Informatik – Organisation, Curricula, Erfahrungen. Bonn: Gesellschaft für Informatik e. V., S. 51-62, 2006.
- [ZS01] Zimmerman, B. J. & Schunk, D. H.: Self-regulated learning and academic achievement: Theoretical perspectives. Routledge, 2001.
- [Zu16] Zukunft, O.: Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen, Bonn: Gesellschaft für Informatik e.V., 2016.

## A Decision Tree Approach for the Classification of Mistakes of Students Learning SQL, a case study about SELECT statements

Heide Faeskorn-Woyke<sup>1</sup>, Birgit.Bertelsmeier<sup>2</sup> and Jan Strohschein<sup>3</sup>

**Abstract:** TH Köln provides a web-based e-learning platform edb<sup>4</sup>, where novices can do their first steps in SQL. The goal of this paper is to build a decision tree (manually) that classifies the novice's errors. To do so we logged data containing tasks, solutions, and wrong statements over seven months and got a table with 7533 rows as a training set. Each leaf node of the decision tree is a class of errors of similar type and generates an error message with feedback to help the user to solve the task. Interesting and surprising are the mistakes that SQL novices make. The result improves the first steps of learning SQL in a simple and personalized way and gives the teachers hints to improve their learning outputs.

**Keywords:** Databases, SQL, web-based learning, e-learning, educational data mining, learning analytics, decision trees.

### 1 Introduction

A lot of different approaches to include e-learning components to a standard database lecture have been used in the last years. A short overview is given in Section 2 of this paper. Here, we report on our development, a decision tree approach to classify the mistakes that SQL novices make to generate hints helping them to avoid these mistakes. We include these hints in the SQL-Trainer of our e-learning platform. We collected 7353 wrong statements over seven months in 2019/2020 and compared them with the correct statement to answer SQL SELECT questions. We use our edb web application mainly for database courses at the Computer Science Institute of TH Köln in the third semester consisting of roughly 300 students. However, the platform is open for other universities too. Five of the total two hundred fifty different SQL tasks are randomly selected in one session, executed, and logged anonymously when the user starts the SQL trainer in our database e-learning platform.

---

<sup>1</sup>TH Köln/Fakultät 10/Institut für Informatik, Gustav-Heinemann-Ufer 54, 50968 Köln, [heide.faeskorn-woyke@th-koeln.de](mailto:heide.faeskorn-woyke@th-koeln.de)

<sup>2</sup> TH Köln/Fakultät 10/Institut für Informatik, Gustav-Heinemann-Ufer 54, 50968 Köln, [birgit.bertelsmeier@th-koeln.de](mailto:birgit.bertelsmeier@th-koeln.de)

<sup>3</sup>TH Köln/Fakultät 10/Institut für Informatik, Gustav-Heinemann-Ufer 54, 50968 Köln, [jan.strohschein@th-koeln.de](mailto:jan.strohschein@th-koeln.de)

<sup>4</sup> See <https://edb2.gm.th-koeln.de>, visited: 25.03.20

The following research questions are covered in this paper:

- Q1:** How can we construct a decision tree to classify SQL errors that students make in their first SQL steps?
- Q2:** How can we derive useful hints and practice to avoid these errors?
- Q3:** What are the main mistakes that students make?
- Q4:** What are the main tips for teachers in SQL that they can apply to improve their learning outputs and decrease dropouts?

## 2 Related Works & History

Automatic SQL e-learning applications are already in use over a long time. A good recent overview of related works can be found in Kleerekoper et al. [KS18]. They treated the automatic assessment of SQL and compared the results that students can reach by this method. Ahadi et al. [Ah15, Ah16a, Ah16b] categorized SQL statements and examined the errors the students make in these particular categories. They used the AsseSQL-Tool of Prior [Pr03], similar to Sadiq [Sa04]. Mitrovic et al. implemented the first constraint-based-tutor in 1998 and published a lot of similar works over time described in [MO16]. Brusilovsky et al. [Br10] combined several e-learning tools to an open architecture. Brass and Goldberg [BG05] made a classification of logical SQL errors with the aspect of logical and semantic error type. Taipalus et al. [TSV18] completed the approach of [BG05], considered an extensive system of logical errors and also complications. Cembalo et al. [CDU11] visualized the intermediate steps of a SQL query up to the final result. Saatz [Sa17] used an NLP approach to compare stored solutions and student solutions by the Levenshtein-Distance. The RelaX tool of the University of Innsbruck [KTS19] transformed SQL Queries into interactive operator trees, and vice versa, operator trees can be transformed into SQL queries. The e-learning databases portal edb<sup>5</sup> exists since 2008, see [Fa13]. The new version has a SQL trainer, a trainer for normalization of relations, a trainer for SQL triggers, procedures and functions, and multiple-choice questions [RF19, Ra09]. Technically the Java Programming Language was replaced by the JavaScript library node.js, additionally using PHP and an Oracle Database in the background. It is a public web application, and registration is possible with any valid email address.

## 3 METHOD AND DATA

The decision tree was constructed manually by the error examples in the training set based on a student project work [HKM19]. The leaf nodes are placed regarding the

---

<sup>5</sup> See <https://edb2.gm.th-koeln.de/>, visited: 25.03.2020



execution order of the clauses of a select statement (i.e. FROM clause with table list before SELECT list of columns and WHERE clause), the difficulty of the corresponding SELECT type or a significant amount of errors in the training data set. The decision tree is implemented as an Oracle PL/SQL function using regular expressions intensively and was tested on our training set. In the leaves, we have error messages and hints for the user to improve his statement. The leaf nodes, which are nearby the root, are executed first. These error messages and resulting hints are implemented in the SQL trainer of our database e-learning platform edb. The first step was to distinguish semantic (43%) and syntax errors (57%). In the background, we used an Oracle database, and so we decided to use the Oracle error messages for syntax errors.

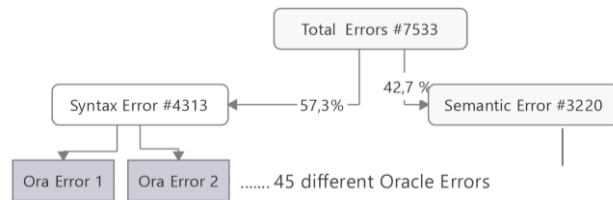


Fig. 1: Syntax errors versus Semantic errors

Compared with Ahadi [Ah16b], we found that the ORACLE syntax errors were more specific than the Postgres errors they used. Most of the errors were misunderstandings of the data model or the correct structure of a simple SQL SELECT statement. Some of the mistakes result from the misunderstanding of the character of the relational database when a set is, for example, compared with a single value (ID 920 and 979).

ID	%	HINT
904	17,8	A column of your query does not exist in the referenced table.
936	11,3	Fundamental syntax error of a select-statement
942	5,7	One of the specified tables or views does not exist.
918	5,5	Please define your columns uniquely by using a table alias!
920	3,7	Incorrect use of comparison operators (<, > =, is) or logical operators (or, and, not)
923	2,9	Missing from-clause
979	2,2	All columns in the select list that do not contain group functions must also be listed in the group by clause.
908	1,1	Operator "is" can only be used with "null"
103	0,1	SQL injection: Other SQL-statements than select are not allowed here

Table 1 Oracle Errors and their frequency of occurring

To distinguish semantic errors, we have two different syntactically correct query results of the wrong statement and the correct statement. Both were executed in the database itself. A user statement is evaluated wrong if the two results differ. As Saatz [Sa17] mentioned, this is only a heuristic method but with 95% accuracy. First, we made a classification of the wrong statements regarding the SELECT-type of the question. Here we expanded the seven SELECT-types of Ahadi [Ah15] to 17 SELECT-types.

Compared with Ahadi [Ah16a], we found similar results. Correlated subqueries, divisions, set operators, recursive SELECTS, WHERE clauses with date column, natural joins, GROUP BY, and ORDER BY clauses have an error quote of larger than 80%. We decided to use the SELECT types with high error quotes and the difficulty of finding errors with text analysis as the first step in our decision tree.

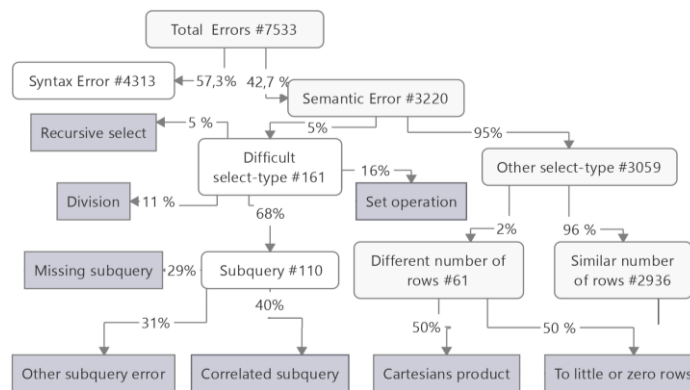


Fig. 2: Decision tree with SELECT types, left side of semantic errors

The next step to construct our decision tree was to make a text analysis with regular expressions to find the errors in comparison to the wrong statement and the correct statement. This is the part under the leaf node "Similar number of rows" in figure 2.

## 4 Results

Regarding our previous questions from section 1, we have the following results:

**Q1:** We used a heuristic and manual manner of constructing such a decision tree. We programmed an Oracle PL/SQL-function that implements the decision tree and classifies all errors of our training set for future use.

**Q2:** From each leaf node of the decision tree, we can derive automatically hints that allow the SQL novice to improve his/her SQL SELECT statements. This is now installed in the public databases e-learning platform edb of our university TH Köln.

**Q3:** We found that most of the errors were due to insufficient knowledge/studying of the data model regarding the table names, the required columns, and the fundamental statement of a SELECT-Statement.

**Q4:** The main problem was, that it is difficult for SQL novices to think in sets. No other programming language is using this kind of rather unusual but straightforward kind of mathematical thinking. Therefore, SQL teachers should spend more time on this type of errors in their lectures to enable student to get more acquainted with this type of structure.

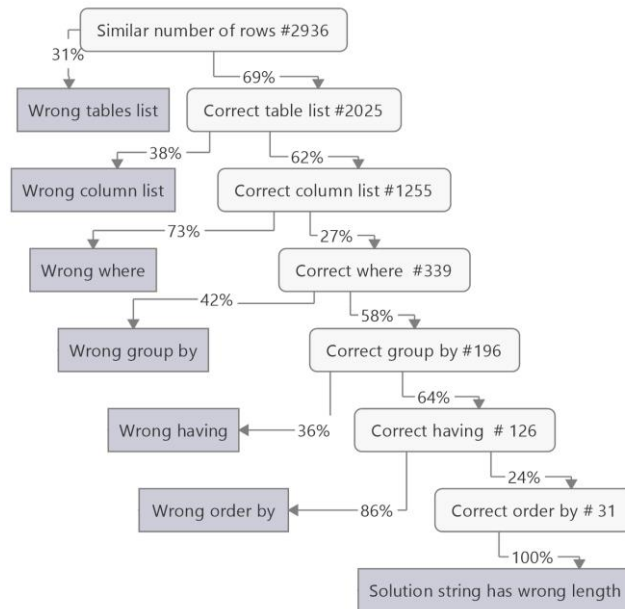


Fig. 3: Decision tree right side for semantic errors

Compared with other works like SQL Tutor of Mitrovic we have a rather simple approach with good results focused on the hints that are generated in our edb platform. In [MO16] about 700 unique constraints were used. In our approach, we needed only 106 leaf nodes and a rather simple program that quickly classifies the wrong solution to an error class and gives the end-user useful hints to avoid those errors.

## 5 Future Work

One possibility for future work is to sample all errors, not only one error, that are concerned with a specific wrong statement. Then the position of the error in the tree does not influence the final result. We will intensively test our algorithm that implements the decision tree, improving the algorithm, and adding more nodes if necessary. For some examples, the oracle errors are not specific enough. We could make a special consideration for those cases, including text analyses. Additionally, we plan to expand our SQL trainer to other SQL DML statements (INSERT, UPDATE and DELETE) as we have done with stored procedures, triggers, and functions and publish the result as an open-source project. A long version of this article and editorial materials like all errors and hints and the decision tree implementation are available on Researchgate<sup>6</sup>.

<sup>6</sup> <https://www.researchgate.net/project/edb-das-ELearning-Datenbankportal>

## Bibliography

- [Ah15] Ahadi, A. et al.: A quantitative study of the relative difficulty for novices of writing seven different types of SQL queries. ITiCSE '15. S. 201–206, 2015.
- [Ah16a] Ahadi, A. et al.: Students semantic mistakes in writing seven different types of SQL queries., ITiCSE'16, S. 272–277, 2016.
- [Ah16b] Ahadi, A. et al.: Students' Syntactic Mistakes in Writing Seven Different Types of SQL Queries and its Application to Predicting Students' Success. SIGCSE 2016, S. 401–406, 2016.
- [BG05] Brass, S.; Goldberg, C.: Semantic errors in SQL queries: A quite complete list. J. Syst. Softw. 79, 5(2005), S. 630–644, 2005.
- [Br10] Brusilovsky, P. et al.: Learning SQL Programming with Interactive Tools. ACM Transactions on Computing Education, 9(4), S. 1–15, 2010.
- [CDU11] Cembalo, M.; De Santis, A.; Umberto, F. P.: SAVI: A new system for advanced SQL visualization. SIGITE'11, S. 165–170, 2011.
- [Fa13] Faeskorn-Woyke, H., Bertelsmeier, B., Gawenda, D. Kasper, A., 2013. Erfahrungen mit dem kooperativen E-Learning-Datenbankportal edb. In: Breiter, A. Rensing, C. (Hrsg.), DeLFI 2013, (S. 285-288).
- [HKM19] Hallweger, M.; Kaczmarczyk, D.; Märtens, F.: Überarbeitung der Fehlerausgabe des SQL Trainers, internal student project, 2019.
- [KTS19] Kessler, J.; Tschuggnall, M.; Specht, G.: RelaX: A webbased execution and learning tool for relational algebra, BTW-GI 2019, S. 503–506, 2019.
- [KS18] Kleerekoper, A.; Schofield, A.: SQL tester: An online SQL assessment tool and its impact. ITiCSE, S. 87–92, 2018.
- [MO16] Mitrovic, A.; Ohlsson, S.: Implementing CBM: SQL-Tutor after Fifteen Years. International Journal of Artificial Intelligence in Education, 26(1), S. 150–159. 2018.
- [Pr03] Prior, J. C.: Online Assesm. of SQL Query Form. Skills. Ace '03, 20, S. 247–256, 2003.
- [Ra09] Rakow, T.C. et al.: Tools für die Lehre im Fach Datenbanken. Datenbank-Spektrum 9. Jg., Nr. 29, S. 5-13, 2009.
- [RF19] Rakow, T. C.; Faeskorn-Woyke, H.: Digitale Lehre im Fach Datenb. In: Meyer, H., Ritter, N., Thor, A., Nicklas, D., Heuer, A. Klettke, BTW 2019 –GI, S. 97-98, 2019.
- [Sa17] Saatz, I. M.: Wo steckt nur der Fehler in der SQL-Anfrage? Semantische Prüfung von Lösungen. CEUR Workshop Proceedings, 2017(Abp), S. 1–9, 2017.
- [Sa04] Sadiq, S. et al.: SQLator-An online SQL learning workbench. SIGCSE Bulletin, 36(3), S. 223–227. 2004.
- [TSV18] Taipalus, T.; Siponen, M.; Vartiainen, T.: Errors and complications in SQL query formulation. ACM Transactions on Computing Education, 18(3), 2018.

## Taking benefit from fellow students code without copying off – making better use of students collective works

Frank Höppner<sup>1</sup>

**Abstract:** We report about research in progress that aims at generating additional value for programming novices from the collective work of all students in this (and earlier) term(s). We propose live progress reports about the whole course to better estimate the own capabilities. And we propose a solution that autonomously creates hints from other students' solutions to support novices in situations where they do not know how to proceed. A plugin integrates the tool seamlessly in the IDE to avoid additional technical burden. We report first results on a small test group.

**Keywords:** programming course, code snapshot, live feedback, formative feedback, hint generation

### 1 Introduction

Introductory programming courses (we consider Java) pose many challenges to novices as they require the development of many skills in parallel. The high number of participants usually prohibits an individual guidance along the way to a solution (e.g. in a programming lab). The advisors are often occupied with approving authorship, while JUnit tests cover functionality (especially in the introductory courses plagiarism checkers may rise too many false alarms for simple exercises). More experienced students are often encouraged to support novices. But true understanding may develop slowly, it needs time. Time is a limited resource for all students, so it seems much more likely that novices are offered a quick repair (“let me quickly fix that for you”) or even a complete solution as an illustrative example (“but do not hand in a 1:1 copy”). This kind of help is well-intentioned, but often a small hint or helpful remark (“you are thinking far too complicated, the solution is much simpler”) might have been enough to bring them back on track. Besides all that, some novices are simply too shy to ask fellow students (let alone advisors) for help.

This work in progress is concerned about utilizing the collective work of a group of students (the current term and/or previous terms) to provide feedback to individual students autonomously. In every course a number of different students solve the very same exercise. Because of the limited complexity of exercises in introductory courses, almost all reasonable solutions should have shown up after a while. So it should be possible to extract additional value out of this resource. We address the following

---

<sup>1</sup> Ostfalia University of Applied Sciences, Dept. Computer Science, Am Exer 2, Wolfenbüttel, Germany, f.hoepner@ostfalia.de

research questions: Is it possible to provide additional value to students in introductory programming courses (Q1) by providing live feedback about the state or progress of the whole group on the current exercise and (Q2) by providing hints about possible next steps by raising questions that were automatically generated from other solutions (but without showing any of the solutions). Of course, we cannot expect hints of the same quality as from experienced advisors, but on the other hand automated hints offer an advantage with respect to availability and anonymity.

## 2 Related work

Data from coding snapshots has been analyzed before. Often the goal is, however, to estimate the student's performance and to predict final exam grades (e.g. [B114]). In contrast, we are not interested in directly estimating the drop-out risk and performance but to provide (automated) feedback (categories taken from the literature review [IV15]).

In [BR06] the authors propose a tool to annotate code with problems or questions, such that a human advisor can provide feedback more efficiently and directly. We aim at something that operates more autonomously. In [An15] reference solutions of the teacher are post-processed by a hint generation tool, which can then reveal parts of the solution successively to the student, depending on their progress. This is related to our intended goal, but again requires manual work and apparently a single reference solution. In our experience, students quite often come up with solutions that deviate somewhat from the intended solution, but are nevertheless valid solutions. It may be annoying to get dragged into a different direction then. Hints were also provided in [DYC10], but the hints directed the students to related tasks and were not obtained by investigating the actual source code, but by finding similarities between the tasks and associations between the covered topics. In contrast, we want to provide hints that are connected to the individual code snapshot and the solution.

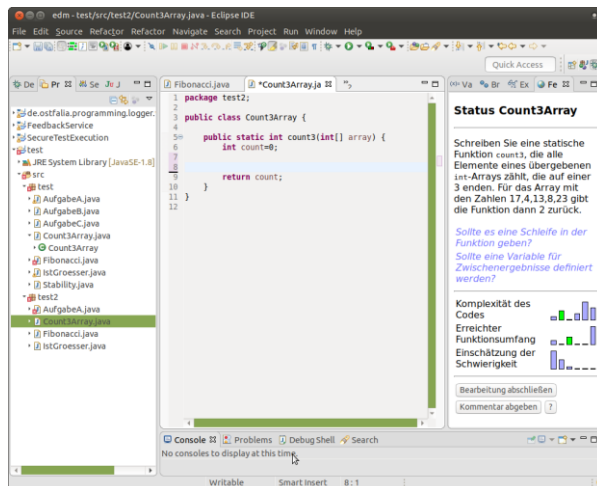
## 3 Proposed Approach

We have implemented a plugin for the Eclipse IDE that submits the student's current code snapshot to a server (not permanently, only if the plugin has been activated for the current project). Once a student has created a Java class with a particular name, which has been associated with a programming task, the exercise instructions appear within the IDE, which makes exercise distribution and handling more convenient. Using a hashcode on the evolution of the code we can keep track of solutions, but neither need the IP addresses nor a user account, so we have no personalized data about the author (privacy preservation).

**Statistics.** With every new submission of the snapshot (every time a student saves the file) some statistics are updated and shown to the student. The currently assessed

statistics address code-complexity (weighted sum of instruction types, where the weight becomes larger for more complex instructions types), perceived difficulty (as voted by students when finally closing the exercise), functional completeness (number of passed JUnit tests), duration since first submission (until completion), and the overall code length. From the set of submissions we construct barplots and present them live in the IDE (right view in Fig. 1). The bar that corresponds to the own solution is highlighted. This allows not only to observe the group of fellow students but to roughly know where one fits in this distribution. Some of these statistics are only shown when the exercise has been completed.

Apart from closing an exercise there is only a single further action the students may invoke: Write an *anonymous message* to point out an unclear part of the task description or request a detailed explanation of some concept during the next lecture.



Translation of feedback window (right): Write a static function `count3` that counts elements of a passed `int`-array that end with a 3. Example: 17,4,13,8,24 should yield 2.

Context sensitive hints: Should there be a loop in the function? Should there be a temporary variable for intermediate results?

Bar charts for code complexity, passed JUnit tests, degree of difficulty.

Fig. 1: Screenshot of the IDE. The view on the right shows the tasks, hints, and feedback statistics.

**Mining Code Snapshots.** The more challenging part is the automatic construction of hints for students who do not know how to proceed or how to attack an exercise. We currently focus on exercises which require the students to write a single function (and test it in the main function). In a nutshell, once a reasonable number of submissions has been received (we used about 20), we (1) extract the core parts of all solutions, (2) group all solutions that follow the same solution path, (3) use the core parts to decide which solution path a partial solution might be following, (4) suggest missing parts as possible next steps. Steps (1) and (2) run offline, (3) and (4) are performed with every new snapshot submission.

To accomplish this, each submission is processed by a parser and the resulting abstract syntax tree is turned into a graph (Fig. 2) by merging nodes that refer to the same variable (e.g. from declarations, expressions or assignments). Every variable usage

therefore establishes a path from the respective instruction to the used variable. Other transformations are applied to the graph, for instance we flatten expression trees to become less sensitive to the exact typing of expressions (e.g. changed evaluation order by using parentheses).

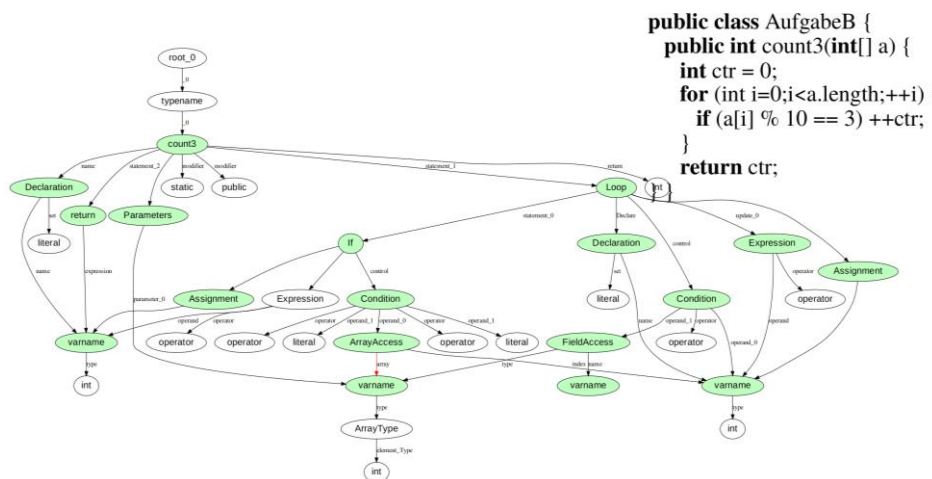


Fig. 2: Graph constructed from top right code snapshot; green nodes match a reference solution

We then identify all *paths* in the graph leading from a variable to the *function declaration* and use them as the core parts of a code snapshot. These paths collectively capture how all variables are used and therefore represent a footprint of the solution. We thus intentionally exclude literals or operators by the aforementioned transformations. These details are relevant to obtain correct results and to pass the tests, but many different expressions may lead to the same results. Including exact expressions in the solution paths would require a much larger collection of code snapshots. So we focus on helping the students to arrive at a valid instruction skeleton.

When comparing a student’s solution against the identified reference solutions (step (3)), we identify these core paths and determine how many paths of a valid solution have already been found. We take these counts as evidence which solution the author may be aiming at. The green nodes in the graph of Fig. 2 show which part of the graph has been matched to a selected instance of reference solutions. Finally, missing paths indicate that certain structural elements are missing in the solution. We transform a missing path into questions in natural language (step (4)), for instance “Should there be a loop in the function?” or “Should there be a temporary variable to store intermediate results”? We phrase the hints as questions rather than giving instructions for two reasons: Firstly, we want the student to think about it rather than blindly following instructions. Secondly, at some occasions the hints may not be correct, because the student follows a different plan than the program is assuming. In the latter case we prefer that the student sticks to her plan rather than feeling advised to change it.



## 4 Evaluation

As this is still work in progress, we can provide a preliminary evaluation only. A group of  $n = 32$  students has tested the plugin and provided feedback by means of a questionnaire that included free text questions as well as expressions of dis/agreement on a Likert-type scale (1=totally agree, 5=totally disagree). The results are shown in Fig. 3.

The students particularly appreciated the tight integration into the IDE (Fig. 3(a)). In the past they all have experienced a somewhat annoying process of submitting solutions to some test facility where the JUnit tests get executed (e.g. copying and pasting it into a learning environment, or submitting it to a version control system). Struggling with the programming language and the IDE already, novices typically do not ask for additional technical burden of such kind. The possibility of asking questions anonymously (Fig. 3(b)) was also highly appreciated (under the assumption that the instructor actually addresses these questions in the lecture, as some students honestly and correctly remarked).

Questions (c) and (d) asked for the utility of the live feedback (live comparison of my own code with that of my fellow students) and the long-term statistics (comparison of my own code against a long-term collection). Overall, the students appreciated both ways of feedback (much more rated  $\leq 3$  than  $\geq 3$ ), but from the written remarks the students were really divided. Some were eager to compare themselves with fellow students in the same course and did not care about former terms (they appreciated a kind of live challenge), others were more reserved because this may give novices a negative feeling of underachievement and instead saw an advantage in comparing themselves to a more representative and larger sample. In a future version, we may let the students decide on their own which type of statistics they prefer to see or provide a stacked barplot with different skill levels. Most of the presented individual statistics were very much appreciated (Fig. 3(f-k)), only the editing time was found to be misleading as it included finished as well as just started exercises at the same time. Here, only the duration of finished exercises should be reported.

The guiding questions for novices were also very much appreciated (Fig. 3(e)). We received remarks such as “The hints make me think about the problem, but do not tell me directly what to do.”, which is very much what we wanted to achieve: We did not want to imitate the IDE’s auto-completion functionality, which can be applied thoughtlessly and therefore does not encourage students to think about the exercise and their code. Others replied that they found the hints encouraging and that they reduced the feeling of being left alone with the tasks. But it needs a more detailed evaluation in the future to figure out to what extent the hints can actually enable students to solve tasks.

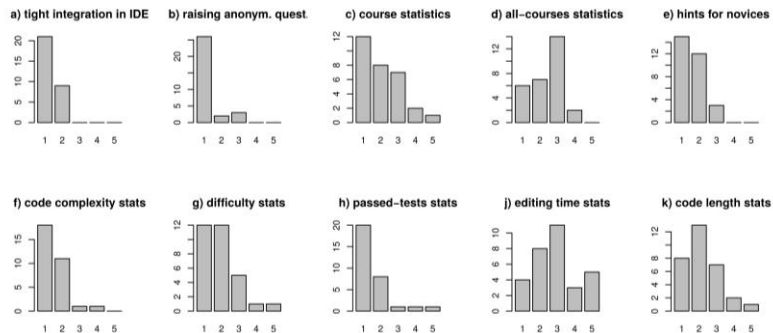


Fig. 3: Answer on a Likert-type scale: Did you find ... helpful (1) or unhelpful (5)?

## 5 Conclusions

Students work alone or in pairs on programming exercises in introductory courses, but usually cannot objectively compare themselves to the course performance. Getting some *live feedback* about the whole course (e.g. “How many tests did others pass by now?”) was considered useful and motivating (which answers **(Q1)** positively), although the respondents were divided whether they want to compare against their own course only or all former courses. Regarding **(Q2)** the fully automatically generated hints, that came in the disguise of questions, were also appreciated and considered helpful for novices who struggle with the next steps of their solution – even though the hints did intentionally not convey too much of the solution. We consider this preliminary evaluation as encouraging and will continue to improve the system and its evaluation.

## Bibliography

- [An15] Antonucci, P. et al.: An incremental hint system for automated programming assignments. In: Conf. Innovation and Technology in Comp. Sc. Education, ITiCSE. pp.320–325, 2015.
- [B114] Blikstein, P. et al.: Programming Pluralism: Using Learning Analytics to Detect Patterns in the Learning of Computer Programming. *Journal of the Learning Sciences*, 23(4):561–599,2014.
- [BR06] Bancroft, P.; Roe, P.: Program annotations: Feedback for students learning to program. In: Australasian Computing Education Conf. pp. 19–23, 2006.
- [DYC10] Dominguez, A. K.; Yacef, K.; Curran, J. R.: Data Mining for Individualised Hints in eLearning. In: EDM. pp. 91–100, 2010.
- [IV15] Ihantola, P.; Vihavainen, A.: Educational data mining and learning analytics in programming: Literature review and case studies. Proc ITiCSE Working Group Report, pp. 41–63, 2015.

## **Maßnahmen zur Verbesserung der Effektivität von *Blended Learning* bei Datenbank-Vorlesungen**

### **Ein langjähriger Erfahrungsbericht**

Richard Lenz<sup>1</sup>, David Haller<sup>2</sup>, Andreas Wahl<sup>3</sup>

**Abstract:** Im vorliegenden Beitrag berichten wir über die Erfahrungen mit Blended Learning und den Effekt einer Probeklausur am Beispiel einer Datenbankvorlesung mit anhaltend hohen Durchfallquoten. Den Studierenden wurde ein breites Spektrum an zusätzlichen digitalen Lernhilfen bereitgestellt. Insbesondere wurde ein semesterbegleitender Online-Selbsttest entwickelt. Die Bereitstellung des digitalen Zusatzangebots alleine hat nicht zur Reduktion der Durchfallquoten geführt. Erst als eine erfolgreiche Teilnahme an den Online-Selbsttests zur Zulassungsvoraussetzung für eine freiwillige Probeklausur gemacht wurde, stieg die Nutzung erheblich und die Durchfallquoten konnten gesenkt werden.

**Keywords:** Blended Learning, Datenbanken, Selbsttests, E-Assessment, Informatikdidaktik, Probeklausur

### **1 Einleitung**

Blended Learning bezeichnet eine hybride Form der Lehre, bei der die klassischen Präsenzveranstaltungen mit computer-basierten dezentralisierten Formen der Lehre kombiniert werden [Gr06]. Im vorliegenden Beitrag berichten wir über die langjährigen Erfahrungen mit einem breit gefächerten Blended-Learning-Angebot zur Ergänzung der Datenbank-Vorlesung „Konzeptionelle Modellierung“ an der Universität Erlangen-Nürnberg. Seit 2011 wird die Vorlesung in jedem Semester hauptsächlich für Studierende der Fachrichtungen Informatik und Wirtschaftsinformatik im ersten Fachsemester angeboten. Dazu kommen Studierende verschiedener Fachrichtungen mit Informatikanteil oder Informatik im Nebenfach. Im Wintersemester gibt es regelmäßig mehr als 500 Prüfungsanmeldungen, im Sommersemester sind es jeweils etwas mehr als 300. Ein Teil dieser Prüfungsanmeldungen sind Wiederholer aus dem Wintersemester.

Mit der Verlagerung der Vorlesung in den Informatik-Studiengängen vom dritten in das erste Fachsemester und mit der Umstellung auf Klausuren mit hohem Multiple-Choice-Anteil sind vermehrt sehr hohe Durchfallquoten im Bereich über 60 Prozent zu beklagen. Im vorliegenden Beitrag berichten wir von einem langjährigen DBR Projekt (Design

---

<sup>1</sup> FAU, Professur für Evolutionäres Datenmanagement, Martensstrasse 3, 91058 Erlangen, richard.lenz@fau.de

<sup>2</sup> FAU, Professur für Evolutionäres Datenmanagement, Martensstrasse 3, 91058 Erlangen, david.haller@fau.de

<sup>3</sup> SIEMENS Mobility, CS DS MAC-DOP, Siemenspromenade 7, 91052 Erlangen, andreaswahl@siemens.com

Based Research), in welchem wiederholt Maßnahmen ergriffen und evaluiert wurden, um diese Durchfallquoten zu senken, ohne den Prüfungsstoff einzuschränken oder das erforderliche Leistungsniveau abzusenken. Basierend auf den Rückmeldungen aus den regelmäßigen Evaluationen der Vorlesungen und Übungen wurden schrittweise neue Zusatzangebote im Sinne des Blended Learning geschaffen, welche die Studierenden in die Lage versetzen sollten, selbstbestimmt die eigenen Defizite schnell zu erkennen und aufzuarbeiten, sowie sich ganz gezielt auf die Anforderungen der Klausur vorzubereiten. Ziel der Bemühungen war eine Verbesserung des Lernerfolgs insbesondere durch die gezielte Unterstützung von Lerninhalten, die oft Schwierigkeiten bereiten. Die Durchfallquoten konnten dadurch nicht substantiell gesenkt werden. Ab dem WS 2018 wurde zusätzlich eine *freiwillige* Probeklausur angeboten, wobei die regelmäßige erfolgreiche Teilnahme an den automatisierten Selbsttests zur Zulassungsvoraussetzung gemacht wurde. Nachfolgend wird berichtet inwiefern durch diese Maßnahme die Durchfallquoten gesenkt werden konnten.

## 2 Verwandte Arbeiten

E-Learning-basierte Werkzeuge zur Unterstützung der Datenbanklehre werden bereits seit vielen Jahren an verschiedenen Hochschulen in Deutschland entwickelt und eingesetzt. (vgl. [Ra09, So10, EJ18, FH13]). Auf besondere Formen des Blended Learning, wie das Modell „Flipped Classroom“ (auch ICM -Inverted Classroom Model) [Di14], oder sogenannte MOOCs (Massive Open Online Courses) [Wi14] soll an dieser Stelle nur verwiesen werden, weil es in diesem Beitrag ausschließlich um freiwillige digitale Zusatzangebote für reguläre Präsenzvorlesungen gehen soll. Untersuchungen zu den Bedingungen, zu denen solche Angebote den Lernerfolg verbessern, zeigen bislang kein klares Bild [vL19]. Zu den speziellen Angeboten zum Thema Datenbanken gibt es zwar Berichte über positive Rückmeldungen von Studierenden, aber ob sich durch die Nutzung der Werkzeuge auch die Klausurergebnisse verbessern ist wenig untersucht. Thomas Rakow zeigt in [Ra09] anhand der Zugriffsstatistik, dass die Datenbank-Lernplattform edb von den Studierenden kontinuierlich gut genutzt wird, wobei vor den Klausurterminen jeweils ein Peak auftritt. Das zeigt zwar, dass das Werkzeug besonders zur Klausurvorbereitung durchaus beliebt ist, es ist aber nicht ersichtlich, ob die Studierenden, welche die Online-Hilfen nutzen, in den Klausuren auch bessere Ergebnisse erzielen.

In der Fachliteratur wurde der Effekt von Probeklausuren sowohl im Hinblick auf ihre Wirksamkeit zur Reduzierung von Prüfungsangst als auch auf ihren Effekt hinsichtlich der Lernleistung untersucht. Jacobs et. al. konnten in einer Studie keinen der gewünschten Effekte nachweisen [JH04]. Dennoch geben in der gleichen Studie 80% der Studierenden an, dass die Probeklausur eine gute Vorbereitung auf die Klausur sei.

### 3 Kombinierte Maßnahmen zum Blended Learning

Die Tabelle in Abb. 1 zeigt das Blended Learning Angebot in der Übersicht sowie den semesterweisen Ausbau. Dabei ging es nicht nur darum, dieselben Inhalte auf verschiedenen Medien zu präsentieren, oder nur verschiedene Zugangsformen zum selben Stoff zu ermöglichen. Stattdessen war das Ziel, einerseits den Studierenden Anreize zu geben, sich mit dem Stoff auseinanderzusetzen, und andererseits flexible Kommunikationskanäle für ein individuelles zeitnahes Feedback einzurichten. In [GK04] wird erläutert, dass gerade diese Mischung aus kognitiven und sozialen Komponenten von zentraler Bedeutung ist.

Verfügbarkeit	Blended-Learning Angebot
Seit Beginn (Sommersemester 2014 und davor)	Alle in der Vorlesung gezeigten Präsentationen (1 Woche vor der Vorlesung).
	Alle Übungsaufgaben (nach der zugehörigen Vorlesung).
	Musterlösungen zu den Übungen (nach der jeweiligen Übung).
	Merkblätter mit Übersichten und zusätzlichen Erklärungen zu schwierigen Themen (mit den zugehörigen Vorlesungsunterlagen).
	<b>Online-Quiz zum Selbsttest</b> (Jedes Online-Quiz wird nach der zugehörigen Präsenzveranstaltung zur Bearbeitung freigeschaltet.)
	Videoaufzeichnung der Vorlesung (Meist kurz nach dem Präsenztermin)
	Zugang zu einer Übungsdatenbank, zur Einübung von SQL. (ständig).
Seit Wintersemester 2015	Online Forum für offene Fragen, die von jedem Kursteilnehmer beantwortet werden können. Bei Bedarf werden Fragen von Tutoren oder dem Dozenten beantwortet. (ständig).
	<b>Kontrollfragen</b> zu jedem Kapitel (Bereitstellung mit den Vorlesungsunterlagen)
	Antworten zu den Kontrollfragen (1 Woche nach der Vorlesung)
Seit Sommersemester 2016	Ausgewählte Klausuraufgaben zusammen mit den zugehörigen Musterlösungen und Erläuterungen. Dabei werden vor allem diejenigen Aufgaben erklärt, die in den Klausuren der letzten Jahre besonders schlecht ausgefallen sind.
	Kontrollfragen pro Kapitel mit zusätzlichen <b>kombinatorischen Verständnisfragen</b> (Bereitstellung mit den Vorlesungsunterlagen)
Seit Wintersemester 2018	Antworten zu den Kontrollfragen (jeweils 1 Woche nach der Vorlesung)
	<b>Online-Quiz ergänzt um z.T. automatisch generierte kombinatorische Verständnisfragen</b>
	<b>Zulassung zur Probeklausur nur in Verbindung mit bestandenen Online-Selbsttests.</b>

Abb. 1: Schrittweiser Ausbau des Blended Learning Angebots

Ein Online-Quiz für den Selbsttest steht seit dem Sommersemester 2012 zur Verfügung. Für diesen Beitrag betrachten wir die Blended-Learning Angebote aller Vorlesungen seit dem Sommersemester 2014, weil die Klausurfragen in diesem Zeitraum zwar variiert wurden, aber in etwa den gleichen Aufgabenmix aufweisen. Das Blended-Learning Angebot wurde gezielt erweitert und auf den Bedarf der Studierenden angepasst. Auf die Fragetypen und Themen, die in der Klausur die meisten Schwierigkeiten verursachen, wurde ein besonderes Augenmerk gelegt. Dazu gehören insbesondere Fragetypen, bei denen das Verständnis von Konzepten überprüft wird, indem Kardinalitäten abgeschätzt werden sollen. Beispiele dafür sind: Maximale Anzahl der Tupel in einer Tabelle, Anzahl der Tupel im Ergebnis einer SQL-Anfrage, Maximalanzahl/Minimalanzahl der Instanzen eines Beziehungstyps in E/R-Diagrammen, usw. Derartige Fragen werden nachfolgend als „kombinatorische Verständnisfragen“ bezeichnet. Auf solche Fragen kann man sich kaum durch Auswendiglernen vorbereiten. Bis zu einem gewissen Grad ist es allerdings möglich, Regeln zur Beantwortung bestimmter Fragetypen zu erlernen. Um diesen Effekt einzugrenzen, werden Fragen und Fragetypen in jeder Klausur variiert. Seit dem Wintersemester 2018 sind auch generierbare kombinatorische Verständnisfragen, die bei jedem Aufruf zu variierten Fragen führen, Bestandteil des Selbsttests.

#### 4 Evaluation zur Effektivität von gemischten Lehrmethoden

Die Durchfallquoten seit dem Sommersemester 2014 sind in Abb. 2 dargestellt.

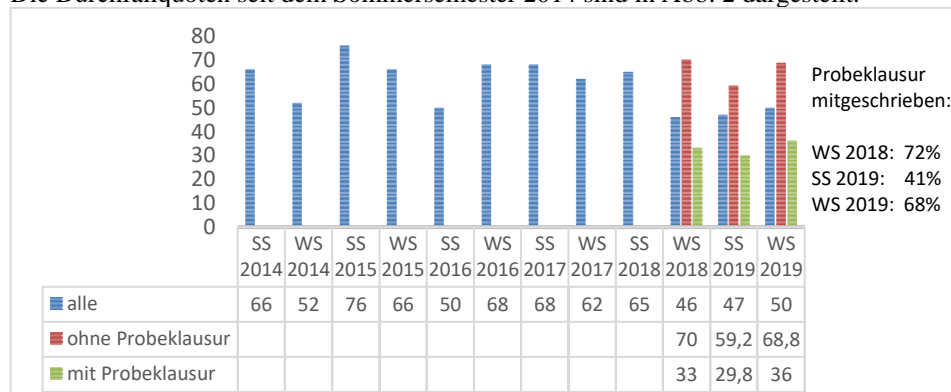


Abb. 2: Durchfallquoten in Prozent der Teilnehmer

Zunächst waren die Durchfallquoten im Sommersemester jeweils regelmäßig höher als im Wintersemester, aber stets über 50%. Dies mag damit erklärt sein, dass die Informatik-Studiengänge die Vorlesung immer im Wintersemester hören. Mit der Bereitstellung der klausurnahen kombinatorischen Verständnisfragen in den Kontrollfragen im Sommersemester 2016 schien der Trend gebrochen, als die Durchfallquote zumindest auf 50% gesenkt werden konnte. Das konnte jedoch bei den nachfolgenden Klausuren nicht

bestätigt werden. In den folgenden Semestern waren die Durchfallquoten stets wieder zwischen 60 und 70%. Die Online-Selbsttests wurden zwar genutzt, aber die meisten Quiz-Teilnehmer haben erst kurz vor der Klausur begonnen, das Angebot wahrzunehmen. Die Selbsttests sind so konfiguriert, dass für jedes Kapitel 80% der erreichbaren Punkte erforderlich sind, um den Test zu bestehen. Viele der Quiz-Teilnehmer haben den Test nicht bestanden. Manche haben dann den Test mehrfach wiederholt, was bei Multiple-Choice Fragen zum Auswendiglernen führt, aber nicht unbedingt zu einem besseren Verständnis. Andere haben nach dem Durchfallen den Test aufgegeben, weil es keine weiteren Tests mit anderen Fragen gab.

Seit dem Wintersemester 2018 wurden die digitalen Lernhilfen zwar immer noch freiwillig, aber nicht mehr unabhängig voneinander bereitgestellt: Um die Lernkontinuität zu verbessern, wurden die Selbsttests so konfiguriert, dass ein Test nur durchgeführt werden darf, wenn der vorherige Test bestanden wurde. Nach erfolgversprechenden Teilnehmerzahlen in den ersten Semesterwochen ging die Nutzung wie in den Jahren zuvor von Woche zu Woche zurück. Um dem entgegenzuwirken, wurde eine freiwillige Probeklausur unter echten Klausurbedingungen angekündigt, die nur mitgeschrieben darf, wenn alle Selbsttests bis zwei Wochen vor der Probeklausur bestanden sind. Als Probeklausur wurde die Klausur aus dem Sommersemester 2018 verwendet. Die Probeklausur wurde nicht korrigiert. Die Teilnehmer erhielten stattdessen eine kommentierte Musterlösung mit Erklärung der Lösungswege. Der Erfolg dieser Maßnahme war deutlich messbar: 72% der späteren Klausurteilnehmer hatten sich für die Probeklausur qualifiziert. Diese Gruppe hatte bei der späteren Klausur eine Durchfallquote von 33%. Unter den übrigen Klausurteilnehmern war die Durchfallquote bei 70%, und damit etwas schlechter aber vergleichbar zu den Vorgängerklausuren. Insgesamt konnte die Durchfallquote auf 46% gesenkt werden. Es ist möglich, dass die Nichtqualifizierung für die Probeklausur oder der Schwierigkeitsgrad der Probeklausur auch Teilnehmer abgeschreckt hat, die eigentliche Klausur mitzuschreiben. Die Teilnahmequoten (Klausurteilnehmer im Verhältnis zu Prüfungsanmeldungen) haben sich allerdings kaum verändert. In den beiden folgenden Semestern wurde das Verfahren noch etwas modifiziert: Um die Lernkontinuität weiter zu verbessern, wurde die Verfügbarkeit der Online-Selbsttests auf zwei Wochen nach der zugehörigen Lehreinheit eingeschränkt. Um zwischenzeitlich erkrankte Studierende dadurch nicht zu benachteiligen, wurde darauf verzichtet, dass ein Selbsttest erst bearbeitet werden kann, wenn der vorherige Test bestanden wurde. Stattdessen wurde festgelegt, dass 8 von 10 Selbsttests bis zu einem Stichtag bestanden sein müssen, um an der Probeklausur teilnehmen zu dürfen. Das Klausurergebnis konnte dadurch aber nicht weiter verbessert werden (vgl. Abb 2). Ob der Effekt sogar negativ war, lässt sich anhand der festgestellten Zahlen nicht bewerten. Immerhin konnte im Sommersemester 2019 die Durchfallquote bei den Teilnehmern der Probeklausur auf 29,8% gesenkt werden. Die Tatsache, dass nur noch 41,8% der Klausurteilnehmer auch an der Probeklausur teilgenommen haben, wird sicherlich zum Teil dadurch erklärt, dass dieselbe Probeklausur wie im Wintersemester angeboten wurde, die von den Wiederholern dann nicht mitgeschrieben wurde. Im Wintersemester 2019 ist wieder die gleiche Probeklausur angeboten worden, zu der sich diesmal 68% der späteren Klausurteilnehmer qualifiziert haben. Insgesamt ist die Klausur geringfügig schlechter

ausgefallen, aber der grundsätzlich positive Effekt der Probeklausur als Anreiz ist weiterhin klar erkennbar. Neben dem Anreizeffekt der Probeklausur konnte die Auswertung der Selbsttests weitere interessante Effekte aufzeigen. So haben beispielsweise Studierende, welche die Tests sehr oft wiederholt haben, im Durchschnitt bei der Klausur sehr schlecht abgeschnitten. Offenbar ist es wichtig sich vor dem Absolvieren der Selbsttests ausreichend mit dem Vorlesungsstoff auseinanderzusetzen.

## 5 Zusammenfassung und Ausblick

Es konnte gezeigt werden, dass eine Probeklausur ein starker Anreiz sein kann, sich kontinuierlich mit dem Vorlesungsstoff auseinanderzusetzen, und dazu auch Online-Selbsttests zu nutzen. Durch die gezielte Abstimmung der Online-Selbsttests auf typische Schwierigkeiten in der Klausur konnten die Durchfallquoten gesenkt werden. Um das Angebot weiter zu verbessern, ist für die kommenden Semester ein zusätzlicher ständig verfügbarer Online-Fragenpool zu allen Themenbereichen geplant, der zur gezielten Nachschulung individueller Defizite genutzt werden kann. Im Sommersemester 2020 steht erstmals auch ein Selbsttest mit frei formulierbaren SQL-Anfragen zur Verfügung. Dazu wurde die Lernplattform ILIAS/StudOn um entsprechende Funktionen erweitert.

## Literaturverzeichnis

- [Di14] Dittrich, J.: Die Umgedrehte Vorlesung – Chancen für die Informatiklehre. Datenbank-Spektrum 14, S. 69–76, 2014.
- [EJ18] El Agha, M. I.; Abdallah M. J.; Samy S. A.: SQL Tutor for Novice Students. 2018.
- [FH13] FH Köln. edb -Das eLearning Datenbank Portal, <http://edb.gm.fh-koeln.de>, 08.08.2020.
- [GK04] Garrison, D. R.; Kanuka, H.: Blended learning: Uncovering its transformative potential in higher education. In: The internet and higher education 7.2, S. 95-105, 2004.
- [Gr06] Graham, C. R.: Blended learning systems. In: The handbook of blended learning: Global perspectives, local designs, S. 3-21, 2006.
- [JH04] Jacobs, B.; Bernd, H.; Fey, A.: Die Wirkung einer Probeklausur auf Klausurleistung und Angst in einer Statistikklausur. Universität des Saarlandes. 2004.
- [Ra09] Rakow, T. C. et al.: Tools für die Lehre im Fach Datenbanken. Datenbank-Spektrum, 9(29), S. 5-13. 2009.
- [So10] Soler, J. et al.: A web-based e-learning tool for UML class diagrams. In: Proc. IEEE EDUCON 2010 Conference, Madrid, S. 973-979, 2010.



## VITMaze – Die Java Coding-Challenge für Verwaltungsinformatiker

Patrick Stalljohann<sup>1</sup>, Maik Merten<sup>1</sup>

**Abstract:** Die Motivation von Studierenden zur aktiven Beschäftigung mit der objektorientierten Programmierung bei sehr heterogenen Vorkenntnissen in interdisziplinären Studiengängen ist eine Herausforderung. Verfügbare Online-Programmierspiele adressieren dies insbesondere für den privaten Bereich, vernachlässigen aber häufig Aspekte der Code-Qualität, der Modularisierung und der Zusammenarbeit im Team. Dieser Erfahrungsbericht zeigt eine Alternative unter Verwendung eines eigenen Programmierspiels auf Basis des CodinGame SDKs auf, das erfolgreich in einem Modul der Verwaltungsinformatik eingesetzt wurde.

**Keywords:** Programmierwettbewerb, Game-Based Learning, CodinGame, Educational Reality, Erfahrungsbericht, Präsenzlehre

### 1 Ausgangslage und Zielsetzung

Im Studiengang Verwaltungsinformatik der Hochschule des Bundes für öffentliche Verwaltung werden Beamtenanwärter auf Ihre späteren Aufgaben im Dienst vorbereitet. Dies erfolgt in derzeit vier parallelen Kursen je Jahrgang mit je ca. 25 Studierenden, welche verschiedene Module zu Themen der Informatik, der Verwaltungslehre und des Verwaltungsrechts absolvieren. Ein Bestandteil ist dabei auch die Vermittlung grundlegender Mechanismen der Anwendungsentwicklung sowie die Befähigung, diese umzusetzen. Dazu zählt insbesondere das Modul „Objektorientierte Programmierung“, in dem entsprechende Ansätze mit der Programmiersprache Java vermittelt werden.

Die Studierenden dieses interdisziplinären Studiengangs sind sehr heterogen in Bezug auf Vorkenntnisse der Programmierung sowie ihren persönlichen Zielen für spätere Aufgaben. Dadurch ergibt sich im genannten Modul die Herausforderung, möglichst viele Studierende zur aktiven Programmierung und tieferen praktischen Umsetzung zu motivieren. Dies wird seit mehreren Jahren mit Hilfe eines Programmierprojektes in Gruppen zum Ende des Moduls nach der Vermittlung der Grundlagen adressiert.

Diese Projekte wurden bis zum Jahr 2018 mit Hilfe der Java-basierten Programmierung von Lego Mindstorms [Le20] durchgeführt. Während damit zwar im Wesentlichen sehr gute Erfahrungen gemacht wurden und dies zur Motivation einiger Studierender beitrug, wurden jedoch auch nachteilige Aspekte beobachtet. So konnten z.B. mechanische Seiteneffekte zu Frustrationen durch unerwartetes Verhalten des Roboters führen. Hinzu

---

<sup>1</sup> Hochschule des Bundes für öffentliche Verwaltung, Fachbereich Finanzen, Verwaltungsinformatik,  
Gescherweg 100, 48161 Münster, {patrick.stalljohann, maik.merten}@vit-bund.de

kommt, dass die Anzahl der parallelen Kurse vorübergehend für einen Jahrgang auf sechs anwächst, um anschließend auf zwei parallele Kurse mit zwei Startterminen im Jahr zurückzufallen. Für diese Spitze sollte die Anschaffung weiterer Roboter vermieden werden, da diese auch nur einmalig genutzt werden würden.

Die Umsetzung einer Alternative sollte daher nicht von zusätzlicher Hardware abhängig sein. Außerdem soll sich jeder der Studierenden in einem Projektteam auch selbständig und ohne Abhängigkeit zur Hardware-Verfügbarkeit mit dem Projekt weiter beschäftigen können. Zu diesem Zweck sollte ein Programmierspiel zum Einsatz kommen, bei dem die Teams spielerisch und kompetitiv motiviert werden, aktiv mit Java zu programmieren.

Nach einer knappen Betrachtung der Problemstellungen von Programmierspielen wird die Programmierumgebung VITMaze vorgestellt, welche ein thematisch zur Verwaltungsinformatik passendes Programmierspiel realisiert. Es folgen genauere Beschreibungen der Spielregeln in mehreren Stufen (Level), eine Erläuterung der Projektphase sowie eine Darstellung des Finales, bei dem die Beiträge aller Teams gegeneinander angetreten sind. Abschließend erfolgen eine Bewertung der Durchführung sowie ein Ausblick auf nächste Schritte und mögliche Anpassungen sowie Erweiterungen.

## 2 Programmierspiele mit CodinGame

Programmierspiele sind computerbasierte Spiele, bei denen die Steuerung des Spielverlaufs nicht direkt durch ein Eingabegerät erfolgt, sondern durch die Ausführung von Programmcode, der von den Spielern eingegeben wird. Sie zählen zur Kategorie der Serious Games.

Eine von mehreren verfügbaren Plattformen ist CodinGame [Co20], auf der zahlreiche Programmierspiele sowohl im Einzelspielermodus als auch in sogenannten Arenen für mehrere Spieler bereitgestellt werden. Dabei wird Programmcode in verschiedensten Programmiersprachen (Java, C#, Python etc.) unterstützt und die Ausführung mit spielerischen Themen grafisch visualisiert. Diese Plattform kann sowohl zum privaten Vergnügen als auch zur Motivation von Studierenden eingesetzt werden [Bu16].

Bei der direkten Verwendung der Plattform durch Studierende, studienbegleitend zur Präsenzlehre, sind bereits positive Erfahrungen bekannt, wobei sich auch Defizite zeigen [HS19]. Hierzu gehört zum Beispiel, dass die Code-Qualität auf der Plattform völlig vernachlässigt werden kann, da allein das Ausführungsergebnis entscheidend ist. Außerdem ist keine Unterstützung von Entwicklungsprozessen im Team vorhanden, da ein ausschließlicher Einzelnutzerbetrieb vorhanden ist. Hierbei erfolgt die Lösung eines Programmierspiels auch grundsätzlich auf Basis einer einzelnen Code-Datei, so dass wichtige Aspekte der Softwarearchitektur, des Entwurfs und der Modularisierung nicht berücksichtigt werden können [HS19]. Mit dem Fokus auf die objektorientierte Programmierung in Gruppen sollte daher eine alternative Verwendung gefunden werden.

### 3 VITMaze

CodinGame unterstützt die Entwicklung neuer Programmierspiele für die Plattform aus der Community durch ein SDK [Co20a]. Der Quellcode des SDKs sowie die Vorlagen für neue Spiele stehen über GitHub zur Verfügung. Kompilierte Versionen können direkt als Pakete aus einem Maven Repository eingebunden werden. Ein neues Spiel wird durch die Implementierung entsprechender Klassen für die Spiellogik eines Spielers (Player) und der Bewertung (Referee) in einer vorgegebenen Vererbungshierarchie realisiert. Testläufe des Spiels können mit der Infrastruktur des SDKs ausgeführt werden, wobei die erzeugte Visualisierung in einer json-Datei aufgezeichnet wird, die in einer enthaltenen Weboberfläche abgespielt werden kann.

Für die Verwendung im Modul zur objektorientierten Programmierung wurde dieser Ansatz genutzt, um das Spiel VITMaze zu realisieren. Dieses Spiel wurde allerdings nicht auf CodinGame bereitgestellt, sondern den Studierenden mit einer Anpassung des SDKs als Ausführungsumgebung zur Verfügung gestellt. Anstatt eines eingeschränkte Online-Editors, sollen eine vollständige IDE verwendet werden können. Dafür wird der Spieler-Code in Form eines Java-Pakets als JAR-Datei bereitgestellt, das mehrere Dateien, Klassen und Referenzen unter einander enthalten kann.

Thematisch sollte das Spiel zur Motivation beitragen und einen gewissen Bezug zur Verwaltungsinformatik besitzen. VITMaze handelt deshalb von der Antragstellung im behördlichen Kontext und Abgabe benötigter Formulare in korrekter Reihenfolge. Spielerisch wurde dies als Verwaltungslabyrinth umgesetzt, in dem sich die Spieler zurechtfinden müssen (siehe Abb. 1). Das zugehörige Spielfeld besteht aus mehreren Feldern unterschiedlicher Art. Das Ziel des Spiels ist es, ein Java-Programm für einen Spieler/Bot zu entwickeln, der sich selbständig in unbekanntem Labyrinth bewegen kann, Formulare einsammelt und diese als erster zum zuständigen Sachbearbeiter bringt.

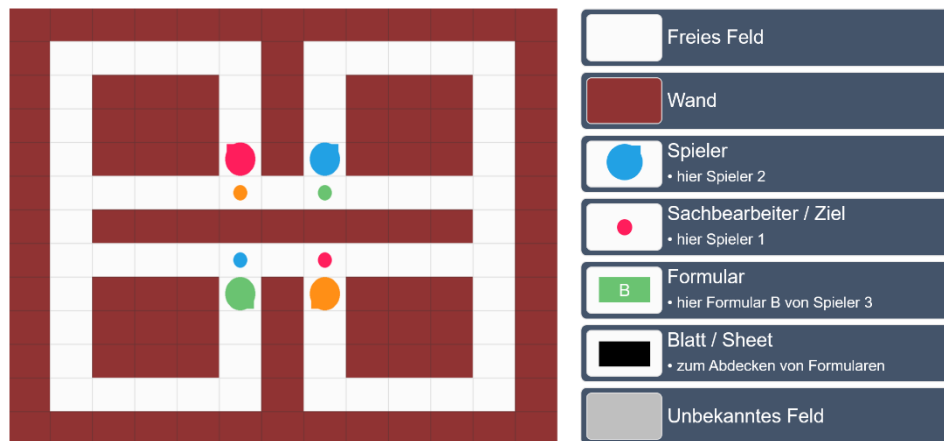


Abb. 1: Beispiel-Spielfeld (Labyrinth) und Legende zu möglichen Feldern.

## 4 Level – Spielregeln und Schnittstellen

Das CodinGame SDK sieht die Standard-Eingabe/Ausgabe als plattformneutrale Schnittstelle vor, auf der ein Textprotokoll zum Austausch von Kommandos und Informationen aufsetzt. Ein Bot liest zunächst die initialen Informationen über die Standardeingabe ein. Anschließend werden die Spielzüge rundenbasiert ausgeführt, so dass jeder Bot weitere Informationen der aktuellen Runde einlesen kann, um daraufhin eine entsprechende Ausgabe zu erzeugen. Diese wird wiederum von der Spielumgebung eingelesen und gemäß den Regeln ausgeführt. Anschließend werden die neuen Ergebniszustände als Informationen der nächsten Runde an die Bots weitergegeben, sofern das Spiel noch nicht beendet wurde. Eine mögliche Kapselung des Protokolls in einen objektorientierten Wrapper wurde bei der Bewertung des Systementwurfs berücksichtigt.

Das Spiel kann in fünf aufeinander aufbauenden Schwierigkeitsstufen (Level) gespielt werden, bei denen die jeweils ergänzenden Spielregeln mit Protokollerweiterungen verbunden sind. Die initialen Informationen für den Bot enthalten die Anzahl der Felder in horizontaler und in vertikaler Richtung, das Level des aktuell ausgeführten Spiels, die jeweilige ID des Bots sowie seine Startkoordinaten. Zu Beginn jeder Runde werden dazu noch das Ergebnis der Aktion aus der Vorrunde sowie der Feldstatus der aktuellen Position sowie des nördlichen, östlichen, südlichen und westlichen Nachbarfeldes bereitgestellt. Die möglichen Status sowie daraus resultierenden Aktionen, die per Ausgabe an die Spielumgebung gemeldet werden können, variieren je nach gespieltem Level.

In der einfachsten Schwierigkeitsstufe (Level 1 – Finden), besteht die einzige Aufgabe des Bots darin, durch das Labyrinth des Spielfeldes zu navigieren. Die Felder mit dem Status `WALL` (Wand) können nicht betreten werden. Die sonstigen Felder sind `FLOOR` (freier Gang) und Sachbearbeiter-Felder (`FINISH <playerId> 0`). Gültige Aktionen je Zug sind die Bewegung in eine Himmelsrichtung (`go north|east|south|west`), die Bestimmung der aktuellen Position (`position`) und das aktive Beenden beim Feld des eigenen Sachbearbeiters (`finish`). Als Ergebnis einer Aktion wird in der Folgerunde ein positiver (`OK`) oder negativer (`NOK`) Aktionsstatus an den Bot übermittelt.

In der nächsten Stufe (Level 2 – Sammeln) müssen vor Beenden der Runde alle benötigten, eigenen Formulare in richtiger Reihenfolge eingesammelt werden. Dafür wird der weitere mögliche Feldstatus `FORM <playerId> <formId>` ergänzt. Eigene Formulare können mit der zusätzlichen Aktion `take` aufgenommen werden, wenn alle Formulare mit geringerer Id auch schon aufgenommen wurden. Die Anzahl der benötigten Formulare ist zu Spielbeginn nicht bekannt, aber nun im Feldstatus des Sachbearbeiters enthalten.

Die Kollision mehrerer Bots in einem Spiel ist ab Level 3 (Unterhalten) möglich. Landen zwei Bots nach einem Zug auf dem gleichen Feld, ist die Aktion im nächsten Zug nicht erfolgreich (`NOK TALKING`) und führt zu Punktabzug. Andere Bots können aber nun auch „gesehen“ werden, indem jedem Feldstatus ein Ausrufezeichen mit Distanzwert nachgestellt ist, falls sich in dieser Richtung ein anderer Bot auf dem Gang befindet.

Aktive Störaktionen sind in den letzten beiden Stufen (Level 4 – Kicken und Level 5 – Verdecken) möglich. Dabei erlaubt die Aktion `kick` zunächst das Verschieben eines gegnerischen Formulars auf ein freies Nachbarfeld. Bei Spielbeginn mit Blättern (Sheets) ausgestattet, können diese mit der Aktion `put` auf ein Spielfeld abgelegt werden und müssen zunächst aufgehoben (`take`) oder weitergeschoben (`kick`) werden, um darunter nach Formularen zu schauen.

## 5 Projektphase

Die erste Projektphase mit VITMaze im Modul zur objektorientierten Programmierung wurde im Jahr 2019 mit 10 Zeiteinheiten à 90 Minuten über einen Zeitraum von drei Wochen durchgeführt. Beides erfolgte in 24 4er-Gruppen, bei denen sich je 2 Studierende frei zu Teams finden konnten, die randomisiert zu 4er-Gruppen zusammengesetzt wurden.

Zunächst wurden das Projekt, dessen Ablauf sowie die Spielregeln von Level 1 vorgestellt. Gefordert wurde außerdem die Anfertigung einer Dokumentation zu Entwurfsentscheidungen, die zusammen mit einem ausführbaren Bot, dem Quellcode und der JavaDoc-Dokumentation bis zum Ende des Projekts eingereicht werden musste. Die Bewertungskriterien zu Funktionalität und Code-Qualität auf deren Basis diese Projektarbeit zu 10% in die Modulnote einging, wurden ebenfalls bekanntgegeben. Der Umgang mit sich ständig ändernden Anforderungen wurde simuliert, indem die erweiterten Regeln und Schnittstellenbeschreibungen der Level 2 bis 5 zu vorher terminierten Zeitpunkten ergänzend bereitgestellt wurden.

Nach Bereitstellung der Ablaufumgebung und Beispiellabyrinthen mit einfachen und komplexeren Gegebenheiten, konnte die Erstellung der Bots in den Gruppen starten. Viele Gruppen haben dazu ein GitHub-Repository angelegt, um gemeinsam und versioniert am Code zu arbeiten. In den Präsenzveranstaltungen wurden die Studierenden bei der Entwicklung betreut und mit Anregungen sowie der Beantwortung von Fragen unterstützt. Neben diesen Präsenzzeiten konnten die Studierenden weiter frei an Ihrer Lösung weiterentwickeln.

## 6 Das Turnier

Zum Abschluss des Projektes traten die 24 Bots gegeneinander an. In jeder Stufe des Turniermodus (Battle) wurden 3 verschiedene Begegnungen (Matches) in unterschiedlichen Labyrinthen gespielt. Zunächst wurden je 4 Bots zufällig zu 6 Start Battles zugeteilt, in den bekannte Labyrinth in Level 2 gespielt wurden. Die beiden besten Bots je Start Battle erreichten die drei Medium Battles, in denen unbekannte Labyrinth in Level 3 anstanden. Die wiederum zwei besten Bots gelangten in die drei Rough Battles in Level 4, bei denen je zwei Spielerbots von zwei Gegnerbots gestört wurden. Die drei Gewinner der Rough Battles gelangten in das gemeinsame Finale, das mit Regeln nach

Level 5 gespielt und durch einen Gegner-Bot gestört wurde. Der Gewinner des Finales wurde zum Siegerbot gekürt und die zugehörige Gruppe erhielt einen Wanderpokal.

Die Durchführung erfolgte nach Abgabe der Bots, so dass die verschiedenen Begegnungen ausgelöst und durchgeführt wurden. Die vorab berechneten Aufzeichnungen der Turnier-Matches wurden in einer Audimax-Großveranstaltung den Teilnehmenden, Lehrenden und Studierenden aller Semester vorgestellt. Hierbei wurde der Turnierverlauf mit einer eigens erstellten Webanwendung visualisiert, so dass "mitgefiebert" werden konnte.

## 7 Feedback, Bewertung und Ausblick

Der sehr positive Eindruck des Autors wurde durch das Feedback der Studierenden im direkten Gespräch, Feedbackrunden und Evaluationsbögen des gesamten Moduls durchgängig bestätigt. Es war zu beobachten, dass die Studierenden sehr viel Zeit mit der aktiven Programmierung des Projektes verbracht haben, auch außerhalb der Präsenzzeiten. Insbesondere wurden aber auch Entwürfe entwickelt sowie geplant und modular programmiert, da von Anfang an klar war, dass Code-Qualität und Struktur in die Bewertung eingehen, das Abschneiden beim Turnier jedoch nicht. Dennoch schien das anstehende Turnier die Motivation vieler Gruppen deutlich zu stärken. Dabei konnten die Studierenden den Umgang mit ihrer Entwicklungsumgebung in der Gruppe (Eclipse + Git) weiter trainieren und auch objektorientierte Konzepte in mehreren Dateien für ihre Abgabe nutzen.

Die größte Anregung der Studierenden war, das Debugging in der laufenden Spielumgebung zu ermöglichen. Aus Sicht des Autors wäre ein Redesign hin zu einer Serverumgebung mit einem service-basierten API wünschenswert, in der der Entwicklungsfortschritt direkt verfolgt werden kann, um eine bessere Transparenz des Lernfortschritts zu ermöglichen.

## Literaturverzeichnis

- [Bu16] Butt, P.: Students' perceptions of game-based learning using CodinGame. In (Morris, L. und Tsolakidis, C., Hrsg.): International Conference on ICT in Education, Rhodes, Greece, S. 151–158, 2016.
- [Co20] CodinGame, Coding Games and Programming Challenges to Code Better, [www.codinggame.com](http://www.codinggame.com), Stand: 18.03.2020.
- [Co20a] CodinGame SDK Documentation, <https://www.codinggame.com/playgrounds/25775>, Stand: 19.03.2020
- [HS19] Hinrichs, T.; Schmolitzky, A.: Einsatz einer Online-Programmierplattform in der Präsenzlehre – ein Erfahrungsbericht. In (Strickroth, S. et al., Hrsg.): 4. Workshop "Automatische Bewertung von Programmieraufgaben" (ABP2019), S. 59–62, 2019.
- [Le20] Lego Mindstorms Education EV3, <https://education.lego.com/de-de/product/mindstorms-ev3>, Stand: 18.03.2020

## Digitale Arbeitsblätter mit interaktiven Programmieraufgaben im Informatik-Unterricht

Sebastian Serth<sup>1</sup>, Ralf Teusner<sup>1</sup> und Christoph Meinel<sup>1</sup>

**Abstract:** Moderner Informatikunterricht umfasst das Erlernen von Grundlagen des Programmierens. Dabei verwenden Lehrer häufig bereits vorhandene Videos, Quizfragen und praktische Programmieraufgaben aus Massive Open Online Courses (MOOCs), obwohl derzeit die Möglichkeiten zur Anpassung der Inhalte und dem Hinzufügen eigener Materialien für Lehrer begrenzt sind. Unsere Software ermöglicht es Lehrern, ihre eigenen interaktiven Arbeitsblätter mit angepassten und eigenen Übungen zu erstellen. Im Rahmen einer praktischen Evaluierung wurde das Konzept von Schülern und Lehrern gleichermaßen gut angenommen: Lehrer hatten mehr Zeit für die Beantwortung individueller Fragen und Schüler konnten in ihrem eigenen Tempo mithilfe automatisierter Rückmeldungen lernen. Für die Vorbereitung zukünftiger Unterrichtsstunden schätzten Lehrer die Möglichkeit, häufige Fehler auszuwerten, um so zuvor unerkannte Probleme besprechen zu können. Interaktive Arbeitsblätter fördern individualisierte Lernprozesse, unterstützen Lehrer in der Unterrichtsgestaltung und sind somit ein wichtiger Bestandteil digitaler Bildung an Schulen.

**Keywords:** Digitale Arbeitsblätter, praktische Programmieraufgaben, Informatik-Unterricht, Schule, MOOC

### 1 Einleitung

In modernem Schulunterricht kombinieren Lehrkräfte unterschiedliche Lernformen und verwenden dazu meist Schulbücher mit Erklärungen und Aufgaben oder stellen eigene Materialien mithilfe von Arbeitsblättern zusammen. In diesem Beitrag konzentrieren wir uns auf die Verwendung von Arbeitsblättern. Einer der größten Vorteile von Arbeitsblättern besteht darin, dass Lehrer die verwendeten Inhalte frei an ihre Bedürfnisse anpassen können. Die vorbereiteten Arbeitsblätter erhalten Schüler dann typischerweise in Papierform. Diese haben allerdings im Vergleich zu digitalen Varianten einige Nachteile, denn gedruckten Arbeitsblättern fehlt jegliche Form der Interaktivität.

Für Informatik-Kurse, insbesondere für solche, in denen Programmierkenntnisse vermittelt werden, ergeben sich zusätzliche Nachteile. Während eine Abbildung, z. B. eine schematische Darstellung eines Herzens im Biologieunterricht, von den Schülern leicht auf Papier annotiert werden kann, ist es schwierig, mit Quellcode auf Papier zu arbeiten. Insbesondere, wenn Schüler den Code ausführen oder verändern müssen, stellt diese Darbietungsform eine unnötige Hürde dar. Ein weiteres Problem tritt auf, wenn Lehrer beschließen, die Programme der Schüler einzusammeln, da diese typischerweise lokal auf dem jeweiligen PC gespeichert sind und die Lehrer somit die Mithilfe der Schüler benötigen.

---

<sup>1</sup> Hasso-Plattner-Institut, Universität Potsdam, Prof.-Dr.-Helmert-Str. 2-3, 14482 Postdam,  
[sebastian.serth@hpi.de](mailto:sebastian.serth@hpi.de), [ralf.teusner@hpi.de](mailto:ralf.teusner@hpi.de), [christoph.meinel@hpi.de](mailto:christoph.meinel@hpi.de)

Ein möglicher Lösungsansatz für die angesprochenen Probleme besteht im Einsatz einer Cloud-Lösung für die Schulen. Eine solche ist dabei die HPI Schul-Cloud [Me19], die spezifisch für das deutsche Bildungssystem entwickelt wird. Derzeit fehlt dem Angebot jedoch eine dedizierte Unterstützung für den Informatikunterricht.

Angesichts der Fülle an Inhalten, die in Massive Open Online Courses (MOOCs) verfügbar sind, nutzen einige Lehrer die angebotenen Ressourcen mit ihren Schülern. Die Online-Kurse enthalten dabei typischerweise Videos und bieten Multiple-Choice-Quizze zur Festigung des Verständnisses oder interaktive Programmieraufgaben an. Die MOOC-Plattformen sind jedoch aktuell nicht auf die Anwendung in der Schule zugeschnitten. Daraus ergeben sich zwei wesentliche Nachteile: (1) Das Kursmaterial ist für Lehrer technisch nicht editierbar, sodass dieses nicht an die Bedürfnisse der Schüler angepasst werden kann und (2) die MOOC-Plattformen genehmigen Lehrkräften in der Regel nicht, Einblick in den Lernfortschritt ihrer Schüler nehmen.

Um diese Nachteile anzugehen, adressieren wir die folgende Forschungsfragestellungen:

1. Wie können wir Lehrer dazu befähigen, vorhandene Aufgaben (z. B. aus MOOCs) wiederzuverwenden, anzupassen und damit eigene, interaktive Arbeitsblätter zu erstellen?
2. Welchen Software-Support benötigen Lehrer, um Schüler zu unterstützen, die Probleme mit den gestellten Programmieraufgaben haben?
3. Wie können die gesammelten Ergebnisse zurück in den MOOC-Kontext fließen, damit eine größere Anzahl an Lernenden ebenfalls davon profitiert?

## 2 Stand der Forschung

Die folgenden Absätze geben einen kurzen Überblick über bisherige Forschungserkenntnisse, auf denen dieser Beitrag aufbaut. Dies umfasst insbesondere den Einsatz digitaler Inhalte im Unterricht und didaktische Ansätze zu Programmieraufgaben im Unterricht:

(1) Interaktive Inhalte im Unterricht: Othman et al. stellten 2013 fest, dass Informatikkurse zahlreiche abstrakte Konzepte enthalten, die sich nicht einfach mit traditionellen Lehrmethoden erklären lassen [OIP13]. Ein Vorteil interaktiver Arbeitsblätter wird von Blayney und Freeman beschrieben: Lernende werden in die Lage versetzt, in ihrer eigenen Geschwindigkeit zu lernen [BF08]. Damit wird die individuelle Gestaltung des Lernprozesses unterstützt, insbesondere dann, wenn ein interaktives Arbeitsblatt verwendet wird, das den Lernenden direktes Feedback gibt und so eine Hilfestellung genau dort bietet, wo Schüler die meiste Unterstützung benötigen. In Bezug auf den Grad der Interaktivität und des Feedbacks haben Blayney und Freeman in einer früheren Arbeit auch gezeigt, dass Multiple-Choice-Quizze nicht ausreichen und häufiges Feedback wichtig ist [BF04].

(2) Programmieraufgaben im Schulkontext: Bei dem Erlernen von Programmiergrundlagen besteht ein wichtiger Teil des Lernprozesses darin, das erworbene Wissen durch das



Schreiben von Quellcode (und dem Üben anderer Kompetenzen, wie z. B. der Modellierung) anzuwenden. Zusätzlich zur Verwendung einfacher Read-Eval-Print-Loops (REPL) wünschen sich Lehrer, dass ihre Schüler kleine Programme entwerfen. Ein guter Weg dazu wird von Isomöttönen et al. beschrieben, die vorschlagen, den Fokus der Schüler mit Quellcode als Vorlage auf bestimmte Themen zu lenken [ILL11]. Dies ist grundsätzlich auch durch Lösungen wie Jupyter (bspw. für Python) möglich, jedoch unterstützen diese die Entwicklung eigenständiger Programme in getrennten Dateien nur eingeschränkt. Plugins für Lernmanagementsysteme (LMS) wie CodeRunner [LH16] fokussieren sich wieder stärker auf die Abgabe einer Lösung und eignen sich nicht gut für deren Erarbeitung.

(3) Einsatz von MOOCs im Unterricht: Die meisten Lehrer, die MOOCs in ihren Unterricht integriert haben, waren mit den Ergebnissen zufrieden [Gr14]. Laut Griffiths et al. ist dies auf die gute Qualität der Inhalte und andere Lehrmethoden als in der Präsenzlehre zurückzuführen [Gr14]. M. Israel zeigte jedoch, dass die Anpassung bestehender Kurse, die nicht auf eine Nutzung in Schulklassen zugeschnitten sind, z. B. in Bezug auf das Engagement der Schüler und die Effektivität des Lernens eine große Herausforderung darstellt [Is15]. Zudem fanden Calfield et al. heraus, dass Lernende aufgetretene Fragen lieber mündlich als über eine Online-Community stellen [CCH13]. Auch Lehrer profitieren von mündlichen Diskussionen, da sie so leichter Einblick in den Fortschritt und Probleme ihrer Schüler bekommen. Dies ist insbesondere von Vorteil, da Lehrer im Allgemeinen keinen direkten Zugang zu den Aktivitäten ihrer Schüler in einem MOOC haben [Is15].

(4) Learning Analytics: Erkenntnisse über das Lernverhalten, so genannte Learning Analytics, können von Lehrern nicht nur zur Leistungsbewertung herangezogen werden, sondern auch für die Verbesserung ihres Unterrichts benutzt werden [MSW14]. Lehrer können die bereitgestellten Informationen zur Vorbereitung zukünftiger Unterrichtsstunden nutzen und so z. B. eine zielgerichtete Wiederholung einbauen, um bestehende Lücken im Verständnis eines Themas zu schließen. In Bezug auf Programmieraufgaben schreibt Blikstein, dass die Häufigkeit der Ausführungen und die Anzahl der Quellcode-Änderungen Rückschlüsse auf die Methodik erlaubt und so einen Einblick in die Herangehensweise und Denkansätze der Lernenden gewährt [Bi11]. Berland et al. betonen, dass oben beschriebene Ergebnisse entweder von einem Lehrer oder für automatisiertes Feedback an die Lernenden verwendet werden können [BBB14].

### 3 Herangehensweise

Im Rahmen dieses Beitrags haben wir qualitative Interviews mit 13 Informatiklehrern, zwei Schulleitern und fünf Schülern geführt. In diesen haben wir Erkenntnisse über häufig verwendete Unterrichtsmethoden gesammelt und mehr über die verfügbaren Ressourcen erfahren. Unsere Interviewpartner hoben dabei insbesondere die Möglichkeiten zur Wiederverwendung bestehender Materialien für die Erstellung von Arbeitsblättern hervor. Aus den Interviews haben wir folgenden Schwächen klassischer Arbeitsblätter unabhängig von der konkreten Verteilungsform (Papier / Word-Dokument / PDF) abgeleitet:

(1) Mangelnde Interaktivität: Einige Lehrer verwenden Beispiele aus Videos oder Online-Spielen wie Pac-Man, um Schüler zu motivieren, an einer bestimmten Aufgabe zu arbeiten (z. B. dem Erarbeiten eines Pfadfindungsalgorithmus für die Geister des Spiels Pac-Man). Ein vorbereiteter Programmrahmen ist dann jedoch von dem gezeigten Beispiel losgelöst und lässt sich durch die Schüler nicht direkt im Rahmen des Arbeitsblatts bearbeiten.

(2) Mangelnde Anpassungsfähigkeit: Andere Lehrer verwenden Inhalte aus MOOCs, z. B. aus den Programmierkursen zu Java und Python, die auf openHPI [MW13] verfügbar sind. In diesen Kursen wird eine browserbasierte Umgebung zum Programmieren und Ausführen von Quellcode namens *CodeOcean* verwendet [St16]. Die Plattform wird von Lehrern als geeignet eingestuft, sie vermissen jedoch eine Bearbeitungsfunktion für Aufgaben. Daher können die Aufgaben für ein Arbeitsblatt nur unverändert mittels Link übernommen werden oder Lehrer müssen einige der Vorteile von *CodeOcean* einbüßen.

(3) Mangelnde Benutzerfreundlichkeit: Durch die digitale Bearbeitung der Programmieraufgaben können Ergebnisse nicht durch Einsammeln von ausgefüllten Arbeitsblättern vollumfänglich gesammelt werden. Einige andere Lehrkräfte kritisierten die Nachteile klassischer Abgabesysteme, die auf dem manuellen Hoch- und Herunterladen von Dateien basieren, sodass diese für den Einsatz im Informatikunterricht impraktikabel erscheinen.

Durch die Analyse mit uns geteilter Arbeitsblätter konnten wir weitere Bedürfnisse ableiten. So verteilen einige Lehrer ihre Arbeitsblätter beispielsweise als bearbeitbare Textdokumente statt als PDF-Dateien, damit Schüler die Dokumente individualisiert ausfüllen können. Andere entscheiden sich für die Nutzung eines LMS, über das Schüler neben der Abgabe von Lösungen auch automatisiert auswertbare Multiple-Choice-Quizze beantworten können. Dennoch fehlt dabei eine Verknüpfung bereitgestellten Quellcodes und anderer Materialien, sodass beim Lernen immer wieder Kontextwechsel erforderlich sind. Wenn Lehrer sich zur Umgehung dieser Nachteile entschließen, z. B. durch die Verwendung von MOOC-Plattformen, müssen sie jedoch Einschränkungen bei der Bearbeitbarkeit der Inhalte und den fehlenden Zugriff auf Lösungen ihrer Schüler akzeptieren. Folglich gibt es derzeit keine Option, die den Bedürfnissen der Lehrer vollständig gerecht wird.

## 4 Konzept der interaktiven Arbeitsblätter

Unser Ziel ist es, Lehrer in die Lage zu versetzen, vorhandene Inhalte in Arbeitsblättern zu editieren und zu erweitern. Die größte Herausforderung für Lehrer stellt dabei die Einbindung und Anpassung von Programmieraufgaben aus MOOCs dar. Um Kontextwechsel zu reduzieren und die Analogie zu einem papierbasierten Arbeitsblatt zu erhöhen zeigen wir die Aufgaben direkt auf derselben Seite wie anderes Material an. Neben formatiertem Text unterstützt unser Arbeitsblatt-Editor auch Bilder und Videos sowie Multiple-Choice-Quizze. Zudem umfasst unser Konzept auch eine tiefgreifende Integration von Learning Analytics. Wie im vorigen Abschnitt erwähnt, stehen diese Daten den Lehrern normalerweise nicht zur Verfügung, obwohl insbesondere sie von speziell aufbereiteten Ansichten

**Vererbung**

Dieses Arbeitsblatt besteht aus 7 Teilen, darunter 3 Programmieraufgaben

**Motivation:**

Ein Zoo hält viele verschiedene Tiere unterschiedlichster Arten. Wenn wir den Zoo mit einer Menge von Klassen und Objekten modellieren möchten, so könnte man für jede Tierart eine eigene Klasse anstellen und pro Tier ein Objekt der jeweiligen Klasse instanzieren. Schauen wir uns dazu ein Beispiel mit Vögeln an ...

```

classDiagram
    class Tier {
        name: String
        order: String
        weight: float
        height: float
        fly: Boolean
    }
    class Woodpecker
    class Tig
    class Pferd
    class Pinguin
    Tier <|-- Woodpecker
    Tier <|-- Tig
    Tier <|-- Pferd
    Tier <|-- Pinguin
  
```

**Video 3.1:**

Vererbung

**Quiz 3.1:**

Wofür wird Vererbung typischerweise verwendet?

- Zur Erstellung von Beziehungen zwischen verschiedenen Klassen.
- Zur Verringerung von Doppelungen von Codezeilen
- Zur Aufteilung der Arbeit am Code auf verschiedene Entwickler.
- Zum Ausdrücken einer „is-a / ist-ein“-Beziehung.
- Zur Darstellung von verschiedenen Generationen bei Lebewesen, also einer Eltern-Kind-Beziehung

**Programmieraufgabe 3.1.2:**

OOP2017 Woche3 Kapitel1 Aufgabe2 (Anzeigen)

```

1- class Story {
2-   public static void main(String[] args) {
3-       DetectiveRobot ronja = new DetectiveRobot();
4-       ronja.spreche();
5-   }
6- }
7- }
8-
  
```

Abb. 1: Beispiel eines Arbeitsblatts zum Thema „Vererbung“, wie wir es in unserer Evaluation verwendet haben. Das Arbeitsblatt besteht aus einer textuellen Erklärung, einem Bild, einem Video, einer Multiple-Choice-Frage und einer Programmieraufgabe.

auf die Daten profitieren könnten. Mit einer detaillierten Analyse jeder Programmieraufgabe soll Lehrern so Einblick darin gewährt werden, wie viel Zeit Schüler für die Bearbeitung einer Übung investiert haben und welche Kommentaranfragen gestellt wurden.

Das vorgeschlagene Konzept zielt darauf ab, Lehrern die Erstellung eigener Arbeitsblätter unter Verwendung anpassbarer Inhalte aus MOOCs zu ermöglichen (siehe Abb. 1). Beim Hinzufügen einer interaktiven Programmieraufgabe sollen Lehrer zudem die Möglichkeit haben, Optionen für die Einbindung einer Aufgabe anzugeben. So wollten manche Lehrkräfte das automatisierte Feedback der Unit-Tests deaktivieren oder die erzielte Punktzahl, z. B. im Rahmen einer Prüfung, ausblenden. Der Arbeitsblatt-Editor muss daher diese Einschränkungen unterstützen und sie beim Aufruf von Programmieraufgaben anwenden.

## 5 Prototypische Umsetzung

Aus technischer Sicht sind drei Komponenten an der Umsetzung unseres Konzepts beteiligt: (1) Die HPI Schul-Cloud als Identitätsanbieter, (2) ein Arbeitsblatteditor namens *edtr.io* und (3) die Ausführungsumgebung *CodeOcean*. Durch das Aufrufen eines Arbeitsblattes wird die Identität des jeweiligen Benutzers und dessen Rolle (Schüler/Lehrer) für eine berechtigungsabhängige Darstellung und das Speichern des Fortschritts verwendet.

*edtr.io* ist ein erweiterbarer, offener Rich-Content-Editor, der Benutzerdaten und Dokumenten in einer relationalen Datenbank speichert und auf diese über GraphQL zugreift. Durch Echtzeitaktualisierungen können Lehrer bestimmte Informationen während der Schulstunde ein- oder ausblenden und so die Aufmerksamkeit ihrer Schüler lenken.

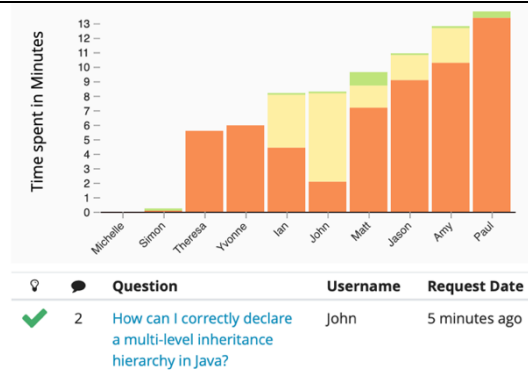


Abb. 2: Das Live-Dashboard stellt die Arbeitszeit dar und schlüsselt über die verschiedenen Farben den aktuellen Fortschritt der Schüler auf: orange (< 50%), gelb (50%-89%), grün ( $\geq$  90%). Außerdem werden unten kürzlich gestellte Kommentaranfragen angezeigt.

Programmieraufgaben können mit wenig Aufwand sowohl umfänglich angepasst (Aufgabenstellung, Testfälle, Vorlage) als auch deren Einbindung samt verfügbarer Features (Code-Editor, Testergebnisse, Hilfestellung, Ansichtsmodus) individualisiert werden.

Beim Aufruf von Programmieraufgaben wird eine pseudonymisierte ID, die für jede Benutzer- und Service-Kombination einmalig ist, die jeweilige Rolle (Schüler/Lehrer) sowie eine Kurs-ID an *CodeOcean* übergeben. Damit werden automatisch Lerngruppen mit allen Mitgliedern der jeweiligen Klasse gebildet werden. In *CodeOcean* können Lehrer zudem detaillierte Informationen ihrer Lerngruppen über ein Dashboard erhalten, das die wichtigsten Aktivitäten pro Aufgabe zusammenfasst (siehe Abb. 2). Es zeigt die von den Schülern investierte Zeit in Korrelation mit den erzielten Punkten und listet Kommentaranfragen, die von den Schülern während der Arbeit an der Aufgabe gestellt wurden.

## 6 Evaluierung

Um sicherzustellen, dass unser Konzept und der entwickelte Prototyp den Bedürfnissen der Nutzer entspricht, haben wir in insgesamt 27 Interviews regelmäßig Feedback zu unterschiedlichen Entwicklungsständen eingeholt.

Um die Auswirkungen unseres Konzepts einschätzen zu können, ist es wichtig zu verstehen, wie Schüler und Lehrer unser Konzept bewerten und wie der Einsatz digitaler Arbeitsblätter den Unterricht verändert. Daher war es das Ziel unserer Evaluierung, möglichst früh Feedback aus typischen Unterrichtsstunden mit unserem Prototypen zu erhalten. Unsere Evaluierung führten wir mit einer Klasse durch, die bereits Java mit Hilfe eines Online-Kurses auf openHPI lernte, um so den generellen Einfluss der Online-Inhalte zu minimieren. In diesem Kurs folgten den Videos Multiple-Choice-Fragen und Programmieraufgaben in *CodeOcean*. Für einen ersten Testlauf bereiteten wir in enger Absprache mit dem Lehrer zwei Arbeitsblätter zu den Java-Themen „Methoden“ und „Vererbung“

vor (siehe Abb. 1). In einem weiteren Durchgang testeten wir ein Arbeitsblatt zu „abstrakten Klassen“, das zudem Referenzen auf das Schulbuch enthielt.

## 6.1 Methodik

Die entworfenen Arbeitsblätter wurden während einer regulären Schulstunde von einem gemischten Informatik-LK aus Baden-Württemberg mit Schülern der 11. und 12. Klasse verwendet. Um eine Beeinflussung unserer Untersuchung durch unsere Anwesenheit auszuschließen, entschieden wir uns gegen eine Hospitation des Unterrichts. Daher lag es in der Verantwortung des Lehrers, das Thema wie üblich vorzustellen und die Schüler durch das Material zu führen. Nach dem Unterricht wurden alle Teilnehmer gebeten, unsere anonyme Umfrage auszufüllen und somit Feedback zu ihren Erfahrungen zu geben. Darüber hinaus haben wir den Lehrer wieder interviewt, um mehr über die Situation in der Klasse und seine Einschätzung zu erfahren. In der Umfrage für die Schüler wollten wir wissen, wie ihnen das Konzept im Allgemeinen gefiel und welche Vor- oder Nachteile sie in dem Einsatz der verschiedenen Medien (d.h. einem Schulbuch, traditionellen Arbeitsblättern und unserem Ansatz) sehen. Zusätzlich erhoben wir den Net Promoter Score (NPS; wie wahrscheinlich es ist, dass sie interaktive Arbeitsblätter an Freunde weiterempfehlen würden), wie er von Reichheld beschrieben wird [Re03]. Zudem haben wir, zur Quantisierung des Nutzungserlebnisses, in unsere Umfrage Teile des Fragebogens zur „modularen Evaluation zentraler Aspekte der User Experience“ (meCUE) [Mi17] eingebunden.

Ferner stellten wir unseren Prototypen interessierten Lehrern auf der didacta, einer der größten europäischen Bildungsmessen, vor. Währenddessen sammelten wir weitergehende Ideen und baten die Lehrkräfte um ihr Feedback zum aktuellen Stand mit Hilfe des standardisierten User Experience Questionnaire (UEQ) [LHS08].

## 6.2 Ergebnisse

Nach der ersten Unterrichtsstunde mit unseren Arbeitsblättern berichtete der Lehrer, dass seine Schüler gut zurechtkamen. Er schätzte die tiefe Integration der verschiedenen Ressourcen sowie den Materialmix, und brachte wiederholt die Absicht zum Ausdruck, selbst gleichwertige Arbeitsblätter erstellen zu wollen. Seiner Beobachtung zufolge wechselten

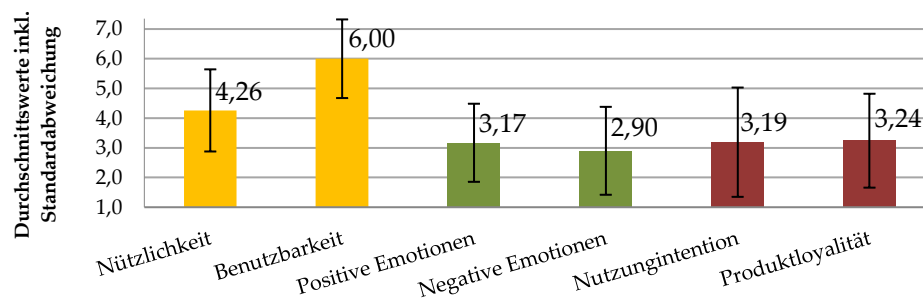


Abb. 3: Ergebnisse der Module I (gelb), III (grün) und IV (rot) des meCUE. (N=7)

die Schüler während der Arbeit an den Programmieraufgaben häufiger zu früheren Teilen der Arbeitsblätter im Vergleich zur bisherigen Lernform mit Online-Kursen.

Von den 21 Schülern der Klasse nahmen neun an unserer freiwilligen Umfrage teil, von denen nur wenige ihre Antworten direkt am Ende der Stunde gaben. Acht Schülerinnen und Schüler schätzten das allgemeine Konzept und das Gesamtdesign unserer interaktiven Arbeitsblätter als gut ein. Sie lobten zudem die gebotene Übersicht über Struktur und Inhalt der Unterrichtsstunde. Außerdem hoben sie die allgemeine Verfügbarkeit der Online-Inhalte (bspw. im Gegensatz zu dateibasierten Arbeitsblättern) hervor, da sie so unabhängig von ihrem jeweiligen Standort auf die Inhalte zugreifen konnten, z. B. auf dem Weg zur Schule. In unserer Umfrage erhielten wir einen NPS von  $-38$  (der zwischen  $-100$  und  $+100$  liegen kann;  $N=8$ ). Die Antworten auf den meCUE zeigten dagegen eine Gesamtpunktzahl von durchschnittlich  $2,4$  (Bereich von  $-5$  bis  $+5$ ;  $N=7$ ). Weitere Werte des meCUE von „stimme stark zu“ =  $7$  bis „stimme stark nicht zu“ =  $1$  sind in Abb. 3 zu sehen.

Von 17 Lehrern, die unseren Prototyp ausprobierten, füllten zwölf den UEQ aus. Die durch den UEQ erfassten Skalen erwiesen sich meist als gut oder ausgezeichnet (vgl. Abb. 4), insbesondere wurde die Stimulation (Mittelwert  $1,7$ ) als am besten bewertet. Im Vergleich zum UEQ-Benchmark, der einen Vergleich mit über  $400$  anderen Studien erlaubt, hat die Durchschaubarkeit ( $1,4$ ) das größte Verbesserungspotenzial.

### 6.3 Diskussion und Interpretation der Ergebnisse

Die Ergebnisse motivieren, den Einfluss interaktiver Arbeitsblätter weiter zu erforschen. Die Umfrageantworten lassen einige, zunächst vorläufige Schlussfolgerungen zu und deuten auf die gute Benutzerbarkeit des Konzepts hin (vgl. meCUE Modul I). Insbesondere die Antworten der Schüler zeigen dabei Widersprüche auf, wie den Unterschied zwischen dem NPS ( $-38$ ) und dem meCUE-Gesamturteil ( $2,4$ ). Auch die Freitextantworten zeigen kleinere Widersprüche, was darauf hindeutet, dass einige Teile unserer Umfrage falsch interpretiert worden sein könnten. Zudem könnte der negative NPS und die neutrale Emotion im meCUE Modul III durch bestätigte Verbindungsprobleme im Schulnetzwerk zum Testzeitpunkt erklärt werden, die auch den Zugriff auf die Umfrage beeinträchtigte. Nach Aussage des beteiligten Lehrers war das Gesamturteil der Schüler ansonsten eher positiv.

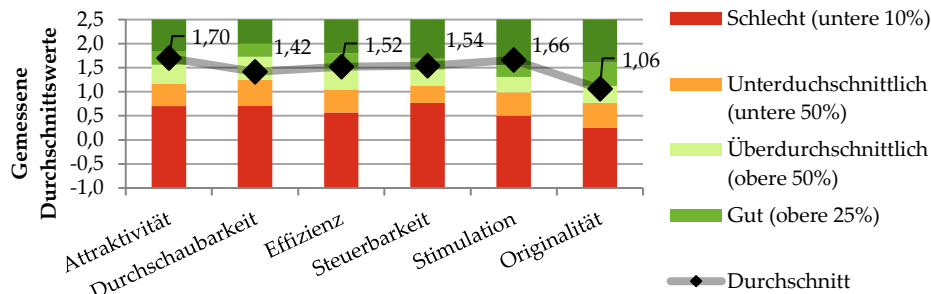


Abb. 4: Durchschnittswerte des UEQ im Vergleich zum Benchmark des Fragebogens. ( $N=12$ )

Aus den Textkommentaren der Schüler ziehen wir drei Schlussfolgerungen: (1) Beim Vergleich der interaktiven Arbeitsblätter mit MOOCs und separaten Arbeitsblättern heben sie die klarere Strukturierung und den verbesserten Überblick der interaktiven Arbeitsblätter hervor. (2) Sie schätzen die Mischung verschiedener Inhaltstypen auf einer einzigen Webseite und sehen keinen Vorteil darin, stattdessen die MOOC-Plattform aufzurufen. (3) Klassische Arbeitsblätter werden als unflexibler angesehen, diese dürfen allerdings manchmal als Nachschlagewerk in Klausuren verwendet werden. Insgesamt schätzen Schüler die Integration der Onlineinhalte und die Einführung moderner Lernmethoden.

Die größten Vorteile interaktiver Arbeitsblätter wurden von Lehrern geäußert. Sie mögen die Einfachheit, mit der individuelle Arbeitsblätter mit Programmieraufgaben erstellt werden können. Unsere qualitative Umfrage zeigt, dass das Konzept den Anforderungen der Informatiklehrer entspricht und die Ergebnisse des UEQ unterstreichen das vorgeschlagene Design. Basierend auf den Erkenntnissen der Learning Analytics gewinnen Lehrkräfte ein besseres Verständnis dafür, warum ihre Schüler mit Aufgaben Schwierigkeiten haben. Die angebotene Darstellung bietet dabei die benötigten Informationen auf einen Blick und wird durch die hinterlegten Unit-Tests automatisch generiert. Während des Unterrichts beobachteten die beteiligten Lehrkräfte, dass die Schüler in ihrem eigenen Tempo zwischen dem Lernen und Anwenden des Stoffs wechselten. Daher berichteten die Lehrer, dass interaktive Arbeitsblätter und das automatisierte Feedback ihnen dabei helfen, von einem Dozenten der Klasse zu einem Lernhelfer für individuelle Schüler zu werden. Das positive Feedback der Lehrer und deren wiederholte Fragen zur Erstellung eigener Arbeitsblätter unterstreicht die Relevanz und praktische Anwendbarkeit unseres Konzepts.

## 7 Ausblick und zukünftiger Forschungsbedarf

Identifizierte Probleme, wie die fehlende Verwendbarkeit der digitalen Arbeitsblätter in Prüfungssituation, sollten mit einer Funktion zum Ausdrucken bearbeiteter Arbeitsblätter behoben werden können. Damit verlieren die Schüler jedoch die Vorteile interaktiver Arbeitsblätter und können während papierbasierter Klausuren ebenfalls nicht auf die gewohnten Funktionen zugreifen. Zukünftige Forschungsarbeiten sollten sich daher auf die speziellen Anforderungen während einer Prüfung konzentrieren. In einem „Prüfungsmodus“ sollten Lehrer die verfügbaren Funktionen noch detaillierter steuern können, z. B. indem ein Zeitlimit vorgegeben wird und Abgaben explizit durchgeführt werden können.

Da die Einführung interaktiver Arbeitsblätter in hohem Maße davon abhängt, wie einfach es für Lehrer ist, eigene Arbeitsblätter zu erstellen, und wie viele Vorlagen bereits zur Verfügung stehen, möchten wir auch Möglichkeiten zum Teilen von Arbeitsblättern und zur Einbindung freier Bildungsressourcen (Open Education Resources, OER) untersuchen. Insbesondere die Erstellung der benötigten Unit-Tests zur automatischen Auswertung erfordert einige Mühen und limitiert Lehrer ansonsten möglicherweise in der Nutzung. Zudem möchten wir unsere Experimente mit einer größeren Nutzergruppe erneut durchführen, um zuverlässige und potenziell verallgemeinerbare Ergebnisse zu erzielen.

## 8 Zusammenfassung

Der Informatikunterricht an Schulen umfasst praktische Programmieraufgaben, durch die technische Anforderungen an Schulcomputer gestellt werden. Daher mögen Lehrer Online-Programmierungsumgebungen (z. B. als Bestandteil von MOOC-Plattformen), die ohne lokale Einrichtung genutzt werden können. Informatiklehrer schätzen die verfügbaren Online-Inhalte und verweisen daher auch von klassischen Arbeitsblättern darauf. Damit Lehrkräfte Inhalte an die jeweiligen Bedürfnisse ihrer Klasse anpassen können, entwickelten wir das Konzept der interaktiven Arbeitsblätter mit Programmieraufgaben. Auf Grundlage unserer Experimente können wir daher unsere Forschungsfragen wie folgt beantworten:

1. Wie können wir Lehrer dazu befähigen, vorhandene Aufgaben (z. B. aus MOOCs) wiederzuverwenden, anzupassen und damit eigene, interaktive Arbeitsblätter zu erstellen?

Lehrer benötigen Zugriff auf die gleichen Werkzeuge wie MOOC-Anbieter und profitieren daher von den neu geschaffenen Möglichkeiten zur Aufgabenbearbeitung in *CodeOcean* sowie der Integration in *edtr.io*. Zudem profitieren sie von Möglichkeiten zum Austausch mit und der Weiterverwendung von Aufgaben ihrer Kollegen. In unserem Konzept können alle Übungen in ein interaktives Arbeitsblatt eingebunden und die den Schülern zur Verfügung stehenden Funktionen mithilfe einiger Parameter nach Belieben angepasst werden.

2. Welchen Software-Support benötigen Lehrer, um Schüler zu unterstützen, die Probleme mit den gestellten Programmieraufgaben haben?

Die Unterstützung durch Software verbessert die Interaktion zwischen Schülern und Lehrern in Bezug auf die folgenden drei Aspekte:

(1) Gesamtüberblick und Einblick in den Bearbeitungsfortschritt der Schüler: Zusätzlich zum Zugriff der Lehrer auf die Abgaben der Schüler bei Bedarf profitieren Lehrer von dem entwickelten Dashboard. Durch die dortige Visualisierung der Ergebnisse der Schüler können Lehrer ihre Schüler bei Verständnisschwierigkeiten gezielt unterstützen und kommende Unterrichtsstunden besser vorbereiten.

(2) Just-in-time-Feedback: Die Echtzeitverarbeitung der Lerndaten in unserem Tool empfinden Lehrer im Rahmen der Evaluierung von Lehrern als hilfreich, da so aufkommende Fragen beantwortet und mögliche Missverständnisse direkt erkannt werden konnten.

(3) Inhaltliche Fragen: Während einer Unterrichtsstunde wünschen Lehrer und Schüler keine technische Unterstützung bei ihrer Kommunikation, da sie die direkte, mündliche Kommunikation bei auftretenden Problemen schätzen. Für Hausaufgaben profitieren Schüler jedoch von der Möglichkeit, Fragen an ihre Lehrer zu stellen oder sich mit zeilenweisen Kommentaren in *CodeOcean* gegenseitig zu helfen.

3. Wie können die gesammelten Ergebnisse zurück in den MOOC-Kontext fließen, damit eine größere Anzahl an Lernenden ebenfalls davon profitiert?



Die verbesserte Struktur der Arbeitsblätter im Vergleich zu einem MOOC wurde von den Schülern geschätzt, und unsere Auswertung legt nahe, dass sie die Lernenden beim erneuten Zugriff auf das Lernmaterial während des Programmierens unterstützt. Zusätzlich zur Unterteilung in feste Wochen profitieren MOOCs zudem von einer besseren Gruppierung der Lernmaterialien desselben Themas. Dies verbessert ebenso die Navigation im Inhalt. Darüber hinaus legt unsere Forschung [Se19] nahe, dass das Konzept der Lerngruppen auf MOOCs angewendet werden sollte, damit ein Tutor die Lernenden individuell durch den Kurs führen kann. So könnten die Tutoren die Kontrolle über die Freischaltung zusätzlicher Inhalte oder die Anpassung der Abgabefristen bekommen und auch Zugang zu den Learning Analytic-Daten erhalten, um informierte Entscheidungen zu ermöglichen.

In diesem Beitrag haben wir digitale Arbeitsblätter mit interaktiven Programmieraufgaben vorgestellt und ausgewertet. Durch die Kombination von Text, Videos und interaktiven Inhalten, einschließlich praktischer Programmieraufgaben und Multiple-Choice-Fragen auf einem interaktiven Arbeitsblatt, verbessern wir den Status Quo sowohl für Schüler als auch für Lehrer: Die Schüler können in ihrem eigenen Tempo lernen und Lehrern wird der Wechsel vom Vortragenden zum persönlichen Tutor erleichtert. Zudem können Lehrer ihre Schüler durch weitere analytische Einblicke besser unterstützen. Die erhaltenen Rückmeldungen aus der Praxis sind überwiegend positiv und zeigen damit das enorme Potenzial, das die Einführung digitaler Inhalte im Unterricht bietet.

## 9 Danksagung und sprachlicher Hinweis

Wir bedanken uns bei allen Beteiligten der Studie für die engagierte Unterstützung. Zur besseren Lesbarkeit wird im Beitrag überwiegend das generische Maskulinum verwendet, diese Formulierung soll jedoch ausdrücklich Personen jeden Geschlechts umfassen.

## Literaturverzeichnis

- [BBB14] Berland, M.; Baker, R. S.; Blikstein, P.: Educational Data Mining and Learning Analytics: Applications to Constructionist Research. *Technology, Knowledge and Learning* Bd. 19, S. 205–220, 2014.
- [BF04] Blayney, P.; Freeman, M.: Automated formative feedback and summative assessment using individualised spreadsheet assignments. *Australasian Journal of Educational Technology* Bd. 20, S. 209–231, 2004.
- [BF08] Blayney, P.; Freeman, M.: Individualised interactive formative assessments to promote independent learning. *Journal of Accounting Education* Bd. 26, S. 155–165, 2008.
- [Bi11] Blikstein, P.: Using learning analytics to assess students' behavior in open-ended programming tasks. In (Long, P. et. al., Hrsg.): *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*. Banff, Canada, S. 110, 2011.

- [CCH13] Caulfield, M.; Collier, A.; Halawa, S.: Rethinking Online Community in MOOCs Used for Blended Learning, 06.10.2013, <https://er.educause.edu/articles/2013/10/rethinking-online-community-in-moocs-used-for-blended-learning>, Stand: 03.04.2020.
- [Gr14] Griffiths, R. et. al.: Interactive Online Learning on Campus: Testing MOOCs and Other Platforms in Hybrid Formats in the University System of Maryland. Ithaca S+R, New York, USA, 2014.
- [ILL11] Isomöttönen, V.; Lakanen, A.-J.; Lappalainen, V.: K-12 game programming course concept using textual programming, In (Cortina, T. J. et. al., Hrsg.): ACM Technical Symposium on Computer science education – SIGCSE '11. Dallas, TX, USA, S. 459, 2011.
- [Is15] Israel, M. J.: Effectiveness of Integrating MOOCs in Traditional Classrooms for Undergraduate Students. The International Review of Research in Open and Distributed Learning Bd. 16, S. 102–118, 2015.
- [LHS08] Laugwitz, B.; Held, T.; Schrepp, M.: Construction and Evaluation of a User Experience Questionnaire. In (Holzinger, A., Hrsg.): HCI and Usability for Education and Work Bd. 5298. Springer Berlin Heidelberg, Berlin u.a., S. 63–76, 2008.
- [LH16] Lobb, R.; Harlow, J.: Coderunner: A Tool for Assessing Computer Programming Skills. ACM Inroads Bd. 7 Nr. 1, S. 47–51, 2016.
- [Me19] Meinel, C. et. al.: Die HPI Schul-Cloud: Roll-Out einer Cloud-Architektur für Schulen in Deutschland. Universitätsverlag Potsdam, Potsdam, 2019.
- [Mi17] Minge, M. et. al.: The meCUE Questionnaire: A Modular Tool for Measuring User Experience. In (Soares, M.; Falcão, C.; Ahram, T. Z., Hrsg.): Advances in Ergonomics Modeling, Usability & Special Populations Bd. 486. Springer International Publishing, S. 115–128, 2017.
- [MSW14] Monroy, C.; Snodgrass Rangel, V.; Whitaker, R.: A Strategy for Incorporating Learning Analytics into the Design and Evaluation of a K-12 Science Curriculum. Journal of Learning Analytics Bd. 1, S. 94–125, 2014.
- [MW13] Meinel, C.; Willems, C.: openHPI: the MOOC offer at Hasso Plattner Institute. Universitätsverlag Potsdam, Potsdam, 2013.
- [OIP13] Othman, A.; Impes, A.; Pislaru, C.: Online Interactive Module for Teaching a Computer Programming Course. In (Ciussi, M.; Augier, M., Hrsg.): Proceedings of the 12th European Conference on e-Learning ECEL 2013. 2013.
- [Re03] Reichheld, F. F.: The One Number You Need to Grow. Harvard Business Review, 2003.
- [Se19] Serth, S. et. al.: Evaluating Digital Worksheets with Interactive Programming Exercises for K-12 Education. In 2019 IEEE Frontiers in Education (FIE). Cincinatti, USA, 2019.
- [St16] Staubitz, T. et. al.: CodeOcean - A versatile platform for practical programming exercises in online environments. In 2016 IEEE Global Engineering Education Conference (EDUCON). Abu Dhabi, S. 314–323, 2016.

## Das PECC-Framework

### Gender-Sensibilität und spielerische Programmierung in der informatischen Grundbildung

Bernadette Spieler <sup>1,2</sup> Carina Girvan<sup>3</sup>

**Abstract:** Mädchen und Frauen sind Vorreiter in der Technologieanwendung, wie zum Beispiel in Bezug auf Smartphones oder Apps. Sie sind aber nur unwesentlich an der Technologieentwicklung beteiligt. Die Jobs der Zukunft sind weitreichend digital und vernetzt und die Nachfrage nach Arbeitskräften mit "Computational Thinking Skills" steigt stetig. Mit einem Fokus auf gendersensible Lehre und Spieldesign können vor allem Mädchen für diese Bereiche motiviert und interessiert werden. Dieser Artikel beschreibt den Einsatz eines geschlechtersensiblen pädagogischen Frameworks für die Informatische Grundbildung. Das "Playing, Engagement, Creativity, Creation" (PECC) Framework wurde durch zentrale Forschungen in den Bereichen Gender Studies, Informatik-Didaktik und von Lerntheorien beeinflusst und iterativ im Zuge einer groß angelegten europäischen Studie entwickelt. In der vorliegenden experimentellen Studie wurde dieser Ansatz mit 12 Schüler und Schülerinnen im regulären Informatikunterricht erprobt und Beobachtung anhand einer qualitativen Inhaltsanalyse beschrieben und mit aktueller Literatur verknüpft.

**Keywords:** Kulturelle und soziale Auswirkungen; Spielentwicklung; Verbesserung des Unterrichts; Sekundarbildung; Lehr-/Lernstrategien

## 1 Einleitung

Die Arbeitsplätze der Zukunft werden weitreichend digital und vernetzt sein und Fachkräfte in den Bereichen Softwareentwicklung und Digitalisierung werden weltweit gesucht [Co17], [Eu18]. Jedoch entscheiden sich immer weniger Studierende für ein technisches Studium und verwerfen die Chance auf eine interessante Zukunftsperspektive und gefragte Berufsaussichten. Insbesondere die MINT-Disziplinen (Mathematik, Informatik, Naturwissenschaften und Technik) und hier vor allem das Studium der Informatik sind davon betroffen. Daher liegt der Fokus auf Schulen und der Forderung nach einem einheitlichen Pflichtfach Informatik, mit der Notwendigkeit Kinder frühzeitig bei der Entwicklung von computergestützten Denkprozessen und Problemlösungsfähigkeiten zu unterstützen. Internationale Beispiele sind das „Computing

---

<sup>1</sup> Universität Hildesheim, Abteilung für Informatik Didaktik, Universitätsplatz 1, 31141 Hildesheim, Deutschland, bernadette.spieler@uni-hildesheim.de, <https://orcid.org/0000-0003-2738-019X>

<sup>2</sup> Technische Universität Graz, Institut für Softwaretechnologie, Rechbauerstraße 12, 8010 Graz, Österreich

<sup>3</sup> Cardiff University, School of Social Sciences, King Edward VII Avenue, Cardiff, CF10 3WT, UK, girvanc@cardiff.ac.uk

Curriculum“ in England [Go13], „Digital Technology“ als Pflichtfach in der Grundschule in Australien [AS17], das fächerintegrative Fach „Digitale Grundbildung“ in Österreich [Bu17] oder außerschulische Programme, wie weltweite „CodeClubs“ [Ra19]. Gleichzeitig sind diese Maßnahmen weltweit mit neuen Probleme verknüpft und Schulen haben oft Schwierigkeiten diese Anforderungen umzusetzen. So ist zum Beispiel in Österreich ein Pflichtfach Informatik nur in der neunten Schulstufe einer Allgemeinen Höheren Schule (AHS) verpflichtend [Bu17]. Informatik wird als ein Wahlfach angeboten beziehungsweise fächerintegrativ als Teil der „Digitale Grundbildung“ unterrichtet. In Deutschland ist eine einheitliche Umsetzung aufgrund der großen Unterschiede zwischen den einzelnen Bundesländern und Schultypen nicht gegeben [St10]. Als Beispiel ist in Niedersachsen, Informatik von der 5. bis 10. Schulstufen fakultativ und wird meist im Rahmen einer Arbeitsgemeinschaft (AG) angeboten. Hier soll ab dem Semester 2023/24 Informatik stufenweise ab Schulstufe 10 als Pflichtfach eingeführt werden [Ni19]. Auch hier wird bereits jetzt eine unzureichende Kapazität an Informatiklehrenden gemeldet. Da die Informatik kein Pflichtfach darstellt, interessieren sich nur wenige für eine Informatik Lehramt Ausbildung. Lehrende in Informatik AGs sind meist nicht für dieses Unterrichtsfach ausgebildet, sondern unterrichten diese Fähig- und Fertigkeiten „nebenbei“ und in ihrem eigenen Ermessen [Co17].

Die Unterrepräsentation von Frauen in der Informatik ist eine wichtige Dimension in diesen Diskussionen. Die Arbeitsplätze ins besonders in der Informatik sind eindeutig männlich dominiert [Eu19a]. In Europa sind 83% der Spezialisten, welche in den Informations- und Kommunikationstechnologien (IKT) arbeiten, männlich. Weitere Statistiken bestätigen einen europaweiten Prozentsatz von nur 17% Studentinnen in IKT-Studiengängen [Eu19b]. Forschende haben festgestellt, dass diese geschlechtsspezifischen Unterschiede nicht auf Industrie- oder Universitätsebene beginnen, sondern bereits im Alter von 12 bis 15 Jahren sichtbar sind (z.B. [Ga15], [Be16]). Diese Unterschiede sind oft bedingt durch gesellschaftliche, kulturelle oder anerzogene Vorurteile, die zu einer geringen Motivation oder sogar Selbstzensur der Mädchen für diese Bereiche führen können [Da14]. Die Annahme der Mädchen, dass Informatik nicht mit ihrem Geschlecht vereinbar wäre, wird durch vorherrschende Stereotypen oder fehlende Vorbilder in diesem Bereich verstärkt [Ga14] und so es wird oftmals ein Beruf gewählt der eher ihrer sozialen Geschlechterrolle entspricht [Ma16], [Ch12]. Initiativen für den Informatiklehrplan berücksichtigen diese geschlechtsspezifischen Unterschiede in Bezug auf Interesse, Selbstwirksamkeit oder Zugehörigkeitsgefühl meist nicht [Sp20].

Angesichts der Herausforderungen, mit denen viele Lehrkräfte zu kämpfen haben, um Lernende für die Informatik gleichermaßen zu interessieren und zu motivieren, müssen geschlechtersensible Strategien entwickelt werden, um sie bei der Gestaltung und Umsetzung im Informatikunterricht zu unterstützen. Um dieses Problem entgegenzuwirken, wurde das Framework für "Playing, Engagement, Creativity, Creation" (PECC) entwickelt. Dieses soll Lehrerinnen und Lehrer dabei unterstützen, motivierende, spielerische und geschlechtersensible Informatik-Aktivitäten in ihren Unterricht zu integrieren. Die Ergebnisse veranschaulichen das Stimmungsbild in der Klasse, indem Einblicke in die Erfahrungen und Perspektiven der Lernenden und Lehrperson im Zuge einer experimentellen Studie gegeben wird.

## 2 Literatur

### 2.1 Geschlechtersensible Pädagogik

Während der Lehrplan weitgehend geschlechtsneutral sein mag, können die Lehrer und Lehrerinnen selbst implizite oder explizite geschlechtsspezifische Vorurteile verinnerlicht haben [St18], [Bo16]. Im Jahre 1995 definierte Connell den Begriff "hegemoniale Männlichkeit" und beschrieb damit, dass die Gesellschaft hauptsächlich Männer mit Macht und wirtschaftlicher Leistung verbindet [Co95]. Die „Nerd-Identität“ in der Informatik repräsentieren die Fachkompetenz im Umgang mit Computer als eine Form der männlichen Fähigkeit. Die Literatur zeigt, dass viele Materialien in der Informatik nicht geschlechterneutral sind und eine männlich-zentrierte Repräsentation in Bild und Sprache verwenden [Me17]. Es ist wichtig, ein realistisches Bild von Frauen in der Technik (daher echte Vorbilder) zu zeigen und Frauen sichtbar und hörbar darzustellen [Fo15]. Eine Sprachsensibilität ist vor allem in Sprachen mit geschlechtsspezifische Substantive, wie Deutsch oder Italienisch wichtig.

In der Praxis bedeutet eine geschlechtersensible Pädagogik: Kinder zu inspirieren, ihre eigene Lernsituation zu erforschen und zu gestalten [Cu14], des weiteren: Leistungsbezogenes Lob, Reflexion der gegebenen Aufmerksamkeit, Interventionen in der Projektphasen und Gestaltung "geschlechtsneutraler" Aufgaben [Wi06]. Darüber hinaus sollen Lehrerinnen und Lehrer Diskussionen und Dialoge anregen, die sich auf individuellen Erfahrungen und Verständnis konzentrieren. So können "sichere" Umgebungen geschaffen werden, für jene welche wenig Vorkenntnis in Informatik haben. Des Weiteren haben viele eine unrealistische Vorstellung von technischen Berufen. Dies beeinträchtigt die Selbstwirksamkeit [Al17], [Ma16]. In Österreich und Deutschland fördern die entsprechenden Bundesministerien konkrete Gender-Kompetenzen [Bu20a], [Bu20b] für alle Lehrkräfte. Aktuell werden diese nur als Wahlfächer angeboten.

### 2.2 Konstruktionistisch inspirierte Pädagogik

Bereits im Jahre 1980, beschrieb Seymour Papert die zukünftige wichtige Rolle von Computern in der MINT-Ausbildung und zeigte Möglichkeiten auf, wie diese Inhalte das computerunterstützte Denken („Computational Thinking“) verstärken. Hier soll Wissen durch (soziale) Interaktionen zum Beispiel in Gruppenarbeiten konstruiert und rekonstruiert werden. „CS Unplugged“ Aktivitäten sind eine Möglichkeit grundlegende Prinzipien der Informatik auf spielerische Weise zu vermitteln ohne tatsächlich einen Computer zu benutzen [La08]. Spiele werden als ein effektiverer Ansatz beschrieben, um zum Lernen zu motivieren [Ka15] und meist mit Charakteristiken wie „Wettbewerb“, „Ausprobieren“ und eine Abfolge von Erfolg und Misserfolgen assoziiert. Kinder erlernen somit eine positive Beziehung zu Misserfolgen herzustellen und aus Fehlern zu lernen [Ha13]. Spielerische Lerninterventionen sind aber vor allem dann erfolgreich, wenn der oder die Lernende, sich allgemein durch Spiele intrinsisch motiviert fühlt.

### 3 Aufbau des PECC Frameworks

Das PECC-Framework aus Abb. 1 wurde über die Jahre sukzessiv weiterentwickelt. Von seiner ursprüngliche Konzeption während des No One Left Behind Projektes (NOLB, H2020 Nr. [645215]) als Game-Making Teaching Framework (GMTF) wurde es auf der Grundlage von neuen Daten [Sp18a], [Sp18b] weiterentwickelt. Das NOLB GMTF beinhaltete vorrangig Merkmale der Spielmechanik und -dynamik mit dem Ziel, die "Create@School"- App in den Unterricht zu integrieren [Sp17]. Die Hauptkomponenten des PECC-Frameworks andererseits: 1. Spielen (Play), 2. Engagement, 3. Kreativität (Creativity) und 4. Erstellen (Creation), sind mit geschlechtsrelevanten Aspekten verbunden. Ziel ist es, die Motivation zu fördern, indem der Schwerpunkt auf die für Mädchen ausschlaggebenden intrinsischen Motivatoren für eine spätere Berufswahl gelegt wurde: Zugehörigkeitsgefühl, Interesse, Selbstwirksamkeit und Spaß [Sp20]. Extrinsische Motivatoren werden berücksichtigt durch anzustrebende Ziele, Spielkonzepte, Gruppenarbeiten und die Möglichkeit zur Umsetzung von eigenen Ideen.

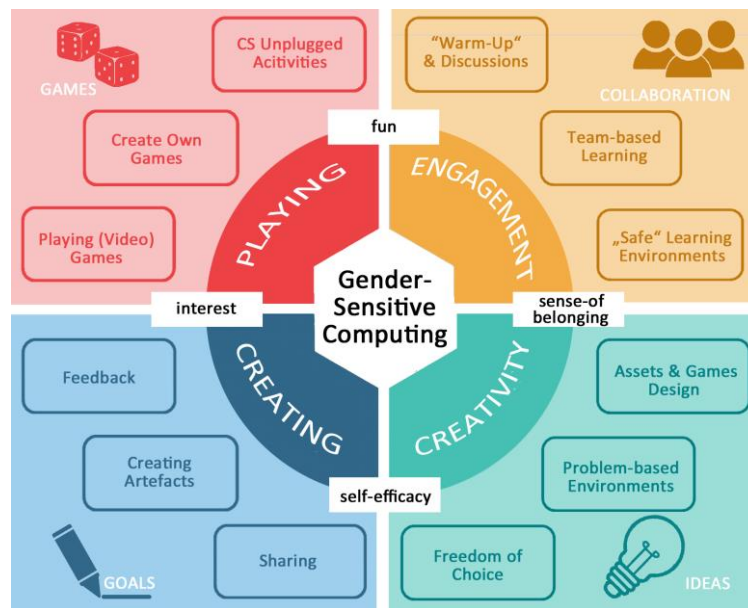


Abb. 1: Das PECC-Framework (Playing, Engagement, Creativity, Creating).

**CS Unplugged:** Erforschen Informatikkonzepte ohne den direkten Einsatz von Computer [Br17]. Verbindung zu realen Problemen schafft ein tieferes Verständnis.

**Eigene Spiele erstellen:** Anleitungen und Vorlagen fördern das Verständnis für Spiele und bieten einen Startpunkt [Ya13], [Vo11]. Spiele selbst zu erstellen unterstützt den Aufbau einer eigenen Wissensstruktur [Ka15].

**(Video-)Spiele spielen:** Berücksichtigen der unterschiedlichen Spielinteressen [He00].

**"Warm-up" & Diskussionen:** zu Berufe in der IT. Echte (weibliche) Role-Models einladen, um die Selbstwirksamkeit zu stärken [Br14].

**Teambasiertes Lernen:** fördert das kritisches Denken, Problemlösung und Kommunikationsfähigkeiten [La18]. Peer-Gruppen Dynamiken oder Rollvergaben innerhalb der Gruppen beachten [Ch15], [Ba16].

**"Sichere" Lernumgebungen:** Zugehörigkeitsgefühl [Ha04]. Manche fühlen sich unwohl Fragen zu stellen [Al17], Wettbewerbe können Ängste auslösen [Gi14].

**Spieldesign:** Selbstwirksamkeit [El07]. Erklärung wichtiger Spiel-Elemente, wie Mechanik (Level, Schwierigkeitsstufen), Dynamik (Punkte, Countdown) und Ästhetik (etwas „Ungewöhnliches“, narrative Elemente) [Hu04] kurz MDA

**Problembasiertes Lernen:** Erforschung verschiedener Inhalte und fördern von kreativen Denkprozessen. Inhalte sind somit nicht isoliert, sondern mit einem gemeinsamen Ziel verbunden, wie der Programmierung eines eigenen Spiels [Pa15].

**Wahlfreiheit:** Ein kreativer Prozess soll die Möglichkeit geben, eigene Ideen zu verwirklichen und die eigene Persönlichkeit widerspiegeln [Sp18].

**Feedback:** Manche brauchen mehr Bestätigung und besonders beim Einsatz von neuen Technologien ist es wichtig, Erfolgserwartungen und den subjektiven Wert der Aufgabe zu erhöhen [Wi06]. Fehler und mehrere Versuche als Teil des Lernprozesses darstellen.

**Artefakte erstellen:** Soziale Relevanz, Geschichten („Storytelling“) und Designaufgaben können ein Teil dieser Aktivität ausmachen [Mc17].

**Teilen:** Stärken des Selbstvertrauen indem fertigen Artefakte freiwillig präsentiert werden oder mit einer Online Community geteilt werden [Br14].

## 4 Forschungsdesign

Eine Gruppe von 12 Schüler und Schülerinnen (5 Jungen, 7 Mädchen, zw. 14-15 Jahre) in einem österreichischen Gymnasium in Graz wurde für drei Wochen im regulären Informatikunterricht begleitet. Für die Erstellung eigener Spiele, wurde die an der Technischen Universität Graz (TU Graz) entwickelte App Pocket Code eingesetzt. Die App bietet eine visuelle Programmiersprache und ist ähnlich der bekannten Scratch-Programmierungsumgebung. Die Lehreinheit wurde gemeinsam mit der Lehrerin nach dem PECC-Framework entwickelt. Die Forschung wurde vor Beginn von der Datenschutzrechtlichen Institution TU Graz und der verantwortlichen Schulbehörde genehmigt und die Zustimmung von Eltern oder Erziehungsberechtigten eingeholt.

Dieser Beitrag konzentriert sich auf die Erfahrungen der Schüler und Schülerinnen und beschreibt diese auf Basis von Notizen, welche während und unmittelbar nach dem Unterricht notiert wurden. Eine qualitative Inhaltsanalyse nach Mayring (2014) dient

dazu, eine ganzheitliche Perspektive des Unterrichts abzubilden und das Framework in einer experimentellen Form zu erforschen. Die beschreibende Analyse der Notizen soll Details zu Gender- und Gruppenaspekten darstellen und zeigen wie diese die PECC-Aktivitäten und Interventionen seitens der Lehrerin beeinflusst haben.

## 5 Ergebnisse & Diskussion

Folgend werden die einzelnen Stufen des Unterrichts beschrieben und mit den Ergebnissen aus den Beobachtungen ergänzt und in Diskussion gestellt. Die Verbindungen zum PECC-Framework sind dabei **fett** dargestellt. Abb. 2 zeigt den Lehrplan des Unterrichts und somit ein Beispiel für die Anwendung des PECC-Frameworks in der Praxis.

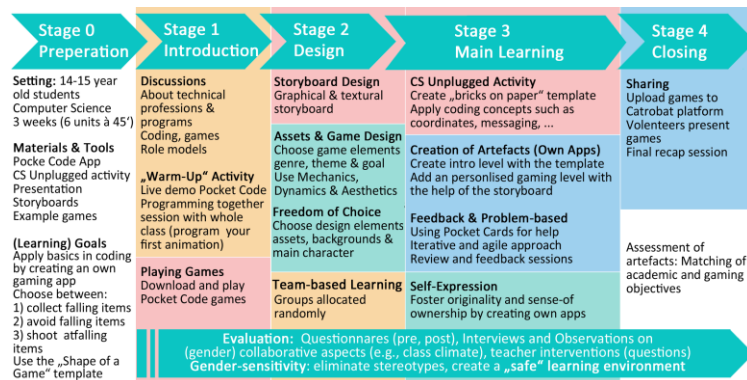


Abb. 2: Ein Beispiel für den Einsatz des PECC-Frameworks im Informatikunterricht.

Stage 0, die Vorbereitungsphase stellt keinen eigenen Bereich im PECC-Framework dar. Diese soll Lehrpersonen helfen, die Struktur vorab festzulegen. Hier muss die geeignete Klassestufe, die Anzahl der Unterrichtsstunden, die verwendeten Technologien, das benötigte Material und das Lernziel definiert werden. Die Wahl des Lernziels hinsichtlich der Spielidee kann bereits die Motivation der Schülerinnen und Schüler beeinflussen. Darüber hinaus evaluieren die Lehrkräfte ihre eigenen Lernmaterialien hinsichtlich genderrelevanter Aspekte (Spiele, Präsentationsfolien, Arbeitsblätter, etc.).

In Stage 1, werden **Diskussionen** zu Informatikthemen angestoßen. Anfangs saßen alle an einem großen Tisch; die Mädchen auf der einen und die Jungen auf der anderen Seite. Daher wurde die Diskussionsrunde in zwei heterogenen Kleingruppen durchgeführt. Hier konnte festgestellt werden, dass ein sehr begrenztes Bild über die Informatik vorhanden war. Besprochen wurden relevante Berufe in der Informatik, eine mögliche Definition von Programmieren, ihre **Spielpräferenzen** (einschließlich bevorzugter Spieleapps) oder berufsrelevante IT-Fähigkeiten. Die Rolle der Lehrerin bestand darin, darauf zu achten wer sich einbringt, wer Fragen stellt, wer die Diskussion eventuell leitet und wer sich zurückhält. Bei Erklärungen, ist es entscheidend, auf eine **geschlechtergerechte Sprache** zu achten und eine hegemoniale Männlichkeit nicht zu fördern. Zum Beispiel auch



einzugreifen, wenn Schüler und Schülerinnen selbst nur vom Programmierer oder Techniker reden. So kann die Geschichte der ersten Programmiererin Ada Lovelace erzählt oder weibliche CEOs wie Virginia Rometty von IBM erwähnt werden. Benannte technische Berufe waren Informatiker/in oder allgemein, „etwas mit einem Computer machen“. Bekannte Aufgaben wurden von den Jungen mit "Programmieren von Anwendungen und Programmen" beantwortet. Ergänzend besagt eine Studie (363 Jungen, 423 Mädchen, 50 keine Angaben), dass selbst wenn die Schülerinnen und Schüler keine Ahnung hatten was Personen in der Informatik tatsächlich tun oder was Programmieren ist, haben 11% angegeben, dass sie aufgrund des Programmierens nicht das Hauptfach Informatik wählen [Ca06]. Das zeigt, wie wichtig es ist ein **realistisches Bild der Informatik** früh genug zu forcieren. Zu Spielen nannten alle Jungen Computerspiele (z.B. FIFA, Sonic) und die Mädchen eher Spieleapps (z.B. Candy Clash, Quiz Duell). Die Jungen lachten über die Antworten der Mädchen und waren unkonzentriert. Die Autorin dieses Berichtes, erzählte als **Role-Model** von ihrem Tätigkeitsbereich und die Lehrerin musste die Jungen mehrmals ermahnen, still zu sein. Als "**Warm-Up**" wurde mit der ganzen Klasse ein kleines Spiel mit den wichtigsten Funktionalitäten in Poket Code erstellt. Dafür diktierten Schülerinnen und Schüler verschiedene Lösungsansätze. Ein Mädchen sagte: "Wir brauchen eine Schleife" (für die Änderung des Aussehens) und ein anderes Mädchen meinte, dass das Objekt den Namen "Hans" tragen sollte. Ein Einwurf von einem Jungen lautete: "Die Nachricht sollte am Anfang gesendet werden". Alle Kinder beteiligten sich aktiv, als die Lehrerin sie nach Fakten zu Planeten befragten (diese benötigten sie für das Spiel-Intro). Im Anschluss an diese Übung, war eine **Spiele-Phase** geplant, wo Spiele mit Pocket Code erstellt heruntergeladen und ausprobiert werden konnten. Diese Möglichkeit wurde von allen genutzt und als sehr positiv gesehen.

In Stage 2 starteten die Schülerinnen und Schüler mit dem **Game Designs** ihrer eigenen Spiele. Dafür wurden zwei Arten von **Vorlagen** verwendet. Eine grafische Vorlage oder Storyboard (Einteilung in Titel, Einleitung, Level und Ende-Bildschirm). Diese wurde nur von zwei der Mädchen verwendet, um eine Skizze des Titels und des Einleitung zu zeichnen. Die textuelle Vorgabe (Titel des Spieles, Genre, Definition der MDAs), war verbindlich und wurde in zirka zehn Minuten von allen ausgefüllt. Um einen Rahmen für ihr Spiel zu definieren, konnten die Schülerinnen und Schüler zwischen drei **Spielzielen** wählen (Ausweichen, Fangen, Schießen, siehe Abb. 2). Zwei Mädchen waren mit den vordefinierten Zielen nicht zufrieden und wollten anstatt dessen ein Puzzlespiel erstellen. Dies wurde von der Lehrerin erlaubt. Die Jungen hatten Spaß daran, lustige Namen für ihre Spiele zu wählen (z.B. "badadibadabdum"). Um **Gruppenarbeiten** zu ermöglichen, wurden die Kinder zufällig durch Ziehen eines Sensors zugeordnet: Berührungsposition, daher ihr Objekt folgt der Fingerposition oder die Steuerung des Objekts mit Hilfe des Neigungssensors des mobilen Geräts. Als die Lehrerin die Schülerinnen und Schüler bat, sich nach dem Sensor aufzuteilen, waren diese zunächst interessiert daran, wer den gleichen Sensor besitzt. Der Sensor wurde jedoch erst zu einem späteren Zeitpunkt benötigt, was zur Folge hatte, dass alle zunächst mit dem Kind neben ihnen zusammenarbeiteten. Die Schülerinnen und Schüler konnten entscheiden, ob sie die Elemente für Hintergründe und Charaktere selbst zeichnen, im Internet nach Grafiken suchen oder diese aus der Medienbibliothek entnehmen (**Wahlfreiheit**). Letztere bietet

eine Vielzahl von Grafiken und Charakteren zur Auswahl. In dieser Phase arbeiteten die Schülerinnen und Schüler sehr individuell an ihren eigenen **Ideen**.

In **Stage 3**, startete die Klasse mit einer „**CS Unplugged**“ Aktivität. Diese Aktivität stellte den Intro-Level des Spieles dar. Dabei handelte es sich um ein A3 Poster, auf dem die ausgeschnittenen Befehle den entsprechenden Objekten unterordnet wurden, siehe Abb. 3. Die „Story“ war die folgende: "Deine Figur möchte etwas über die Planeten lernen und macht sich auf eine Reise ins Weltall. Die Planeten erzählen Fakten über sich selbst."

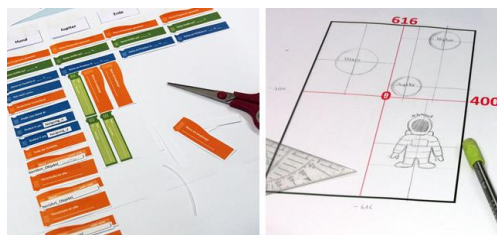


Abb. 3: „CS-Unplugged“-Aktivität. Links: Bausteine werden unter Objekte platziert. Rechts: Objekte werden auf bestimmte Koordinaten „gesetzt“.

Für die „CS-Unplugged“ Aktivität begannen die Schülerinnen und Schüler damit ihre Objekte zunächst auf ein ausgedrucktes Koordinationssystem zu legen oder zu zeichnen und die Befehle einzeln auszuschneiden (siehe Abb. 3). Vier der Mädchen und ein Junge brauchten eine zusätzliche Erklärung zum Koordinationssystem (platzieren der Objekte). Einige fragten, wo und warum bestimmte Befehle benötigt werden, die nicht im Plenum erklärt wurden (2 Jungen, 3 Mädchen), und zwei der Mädchen ignorierten diese Befehle. Zwei der Mädchen schnitten die Befehle sehr genau aus. Die Lehrerin wies sie an, fortzufahren und half ihnen beim Ausschneiden. Im Allgemeinen wurden alle von der Lehrerin ermutigt, sich eher gegenseitig zu helfen anstatt jeden Schritt zu erfragen. Trotzdem warteten drei der Mädchen auf Instruktionen und drei der Mädchen fragten die Lehrerin nach jedem Schritt. Nur ein Mädchen versuchte, die Aktivität selbst zu lösen und arbeitete alleine an der Aufgabenstellung. Die Jungen hingegen arbeiteten alle gemeinsam und versuchten den Zweck der einzelnen Befehle selbst herauszufinden. Einer von ihnen nahm eine leitende Rolle ein und führte die Gruppe der Jungen an. Dieses Phänomen lässt sich auch in anderen Studien aus Informatikklassen beobachten [De15]: Jungen fragen eher ihre männlichen Sitznachbarn und neigen dazu den Code voneinander zu kopieren. Mädchen fragen eher die Lehrperson um Hilfe. Infolgedessen können Lehrpersonen, die diese geschlechterspezifischen Gegebenheiten ignorieren, ein Umfeld schaffen, in dem die Schüler die sachkundigere Position einnehmen und die Schülerinnen sich in der weniger sachkundigen Position wiederfinden. Am Ende waren die Mädchen in der Gruppe jedoch schneller mit dieser Aktivität fertig. Für diese Recherche benutzten alle Smartphones. Während Mädchen nach interessante Fakten zu Planeten suchten, wurden die Jungen durch das Smartphone abgelenkt (installierte Apps).

Für die eigene Programmierung (**Artefakten erstellen**) starteten die Kinder mit einer **Programmiervorlage** mit vorprogrammierten Szenen für Titel, Einleitung und Ende. Diese mussten zuerst mit Inhalten befüllt und verknüpft werden. Als nächstes fügten alle

die Befehle aus ihrer eigenen gebastelten „CS-Unplugged“ Aktivität zum Spiel hinzu. Während dieser Aufgabe arbeiteten alle in **kleinen Gruppen** von zwei bis drei Personen und halfen sich hauptsächlich gegenseitig. Bei vielen funktionierte dieses Intro beim ersten Versuch nicht. Die Lehrerin verwies auf einen iterativen Ansatz: Einen Befehl hinzufügen und testen. Auf diese Weise wird ein agiler Ansatz erlernt (algorithmisches Denken). Die Schüler und Schülerinnen lernten aus den Fehlern der anderen, aber generell konnte beobachtet werden, dass die Mädchen wieder dazu neigten die Lehrerin zu fragen und Jungen sich eher gegenseitig halfen. Nur Zwei (1 Mädchen, 1 Junge) versuchten allein die Lösungen zu finden. Alle bis auf zwei Mädchen beendeten den Introlevel am Ende der zweiten Woche. Einige Bilder zu Planeten waren in der Medienbibliothek nicht vorhanden und wurden vor allem von Schülerinnen selbst erstellt. Auch hier waren die Mädchen beim Zeichnen sehr präzise und investierten viel Zeit in diese Aktivität. Die Lehrerin wies sie an, zuerst die Skripte hinzuzufügen und Planeten später zu vervollständigen. Ein Junge sagte, er habe das Intro abgeschlossen und begann sehr früh mit seinem eigenen Spiel. Es stellte sich später heraus, dass noch Fehler im Programmiercode vorhanden waren. Spätestens in der dritten Woche starteten alle mit ihrem **eigenen Spiel**. Jene die noch nicht fertig waren, wurden von der Lehrerin angewiesen auch zu starten (2 Mädchen). Jede Unterrichtsstunde begann mit einer **Feedbackschleife** und der Definition von **Teilzielen**. Die Lehrkraft **lobte** die Schülerinnen und Schüler und fördert sie zur **Zusammenarbeit** auf und förderte die **Kommunikation** zwischen ihnen. In dieser Phase, blödelten einige der Jungen herum und das Klassenklima war eher laut. Jetzt zogen es auch die Schüler vor, vorwiegend die Lehrperson zu fragen. Ein Grund dafür war, dass alle an eigenen Ideen arbeiteten. Die Rolle der Lehrerin bestand darin, die Schülerinnen und Schüler zu motivieren, indem Verbesserungen und Erweiterungen vorgeschlagen wurden, wie das Hinzufügen von Sounds. In der zweiten Hälfte der Unterrichtsstunde konzentrierten sich die Jungen im Gegensatz zu den Mädchen mehr darauf ihre Spiele fertigzustellen. Es schien als wären die Schüler eher aufgabenorientiert als intrinsisch motiviert. Die Mädchen hingegen wollten "schöne" Projekte erschaffen, verbrachten teilweise viel Zeit mit dem Zeichnen einzelnen Objekten und weniger mit der Programmierung, waren aber auch sichtlich stolz auf ihre Projekte. Am Beginn der letzten Stunde, war das Spiel eines Mädchens und eines Jungen fast fertig. Dies schien für die anderen Jungen sehr motivierend zu sein ("Ich möchte auch ein so cooles Spiel haben"), im Gegensatz zu den Mädchen ("Ich kann nie ein so cooles Spiel programmieren."). Das Mädchen, wurde früher fertig und sie half zwei anderen Mädchen, als die Lehrerin sie darum bat. Die Mädchen fühlten sich im Allgemeinen beim Finalisieren ihrer Spiele mehr gestresst, während die Jungen bereits die Spiele der anderen ausprobierten und sie laut kommentierten (meist negativ: "Haha! Dein Spiel funktioniert nicht!"). Ein Junge war frustriert, weil die anderen ihn hänselten und setzte sich folglich an einen anderen Tisch. Jene Jungen, die mit dem Spiel fertig waren, waren weniger motiviert weitere Funktionen hinzuzufügen und begannen auf dem Smartphone installierte Apps zu spielen.

In Stage 4 wurden alle Projekte auf die Catrobat Community-Seite geteilt (**Teilen**). Ein Junge und ein Mädchen stellten ihre Spiele freiwillig vor und die anderen wurden ermutigt Fragen zu stellen (**Feedback**).

## 6 Conclusio

Dieser Beitrag beschreibt den Einsatz des PECC Frameworks in der Praxis mit Hilfe einer deskriptiven Analyse. Dabei werden Aktivitäten in den verschiedenen „Stages“ und die Dynamik in der Klasse beschrieben. Das PECC Framework konzentriert sich auf die intrinsische Motivation der Lernenden. Es soll das Zugehörigkeitsgefühl für die Informatik gefördert, das Interesse für diesen Bereich erweckt und die Selbstwirksamkeit für Informatische Themen gestärkt werden. Schließlich sollen sie auch Spaß an den Aktivitäten haben. Die extrinsischen Motivatoren des PECC Frameworks bieten definierte (Lern-)Ziele, schlagen Spiele als einen motivierenden Ansatz vor und zeigen wie wichtig Gruppenarbeiten und die Umsetzung eigener Ideen sind. Dadurch bietet PECC nicht nur einen optimalen Ausgangspunkt für alle, sondern fördert auch den Aufbau von Wissen in einem kreativen, geschlechtssensiblen und nicht wettbewerbsorientierten Umfeld. Für diesen Beitrag war es wichtig, ein vollständiges Bild einer PECC-Aktivität aufzuzeigen, die intrinsischen und extrinsischen Motivatoren der Teilnehmenden zu berücksichtigen und sich auf geschlechtsrelevante Beobachtungen zu konzentrieren. Das PECC Frameworks gab dabei den Rahmen vor und half der Lehrerin, geschlechtersensibel zu agieren. Diese Studie erwies sich für eine Weiterentwicklung des Frameworks als sehr hilfreich. In Zukunft soll das PECC Framework auch außerhalb des schulischen Rahmens getestet werden und weitere Forschungen, zum Beispiel in größeren Gruppen oder geschlechterheterogen und homogenen Gruppen (Girls-only) werden durchgeführt.


## Literaturverzeichnis

- [Al17] Alvarado, C.; Cao, Y.; Minnes, M.: Gender Differences in Students' Behaviors in CS Classes throughout the CS Major: Proc. 2017 ACM SIGCSE Technical Symposium on Computer Science Education, S. 27-32, 2017.
- [As17] Assessment and Reporting Authority <http://www.australiancurriculum.edu.au/>, Stand: 21.02.2020, 2017.
- [Be16] Beyer, S.: Women in CS: Deterrents. Encyclopedia of Computer Science and Technology Edition: 2ndChapter: Women in CS: Deterrents, P. A. Laplante, 2016.
- [Br17] Brackmann, C.; Román-González, M.; Robles, G.; Moreno-León, J.: Development of Computational Thinking Skills through Unplugged Activities in Primary School. Proc. 12th Workshop on Primary and Secondary Computing Education, S. 65-72.
- [Bu20a] Bundesministerium für Bildung und Forschung: Genderforschung. <https://www.bmbf.de/de/genderforschung-222.html>, Stand: 02.04.2020.
- [Bu20b] Bundesministerium für Bildung, Wissenschaft und Forschung: <https://www.bmbwf.gv.at/Themen/schule/schulpraxis/uek/gender>,
- [Bu17] Bundesministerium für Bildung, Wissenschaft und Forschung, <https://tinyurl.com/y78wov7a>. Stand: 02.04.2020.
- [Ca06] Carter, L.: Why students with an apparent aptitude for computer science don't choose to major in computer science, SIGCSE Bulletin, S. 27-31, 2006.

- [Ch15] Chance, S.M.; Bowe, B.: Influence of Collaborative Learning on Women's Experiences of Engineering Education: Research in Engineering Education Society, 2015.
- [Ch12] Cheryan, S.: Understanding the paradox in math-related fields: Why do some gender gaps remain while others do not? In Commentary. *Sex Roles*, S. 184-190, 2012.
- [Co17] Committee on European Computing Education. <https://www.informatics-europe.org/component/phocadownload/category/10-reports.html?download=60:cece-report>. Stand: 21.02.2020, 2017.
- [Co95] Connell, R., 1995. *Masculinities*. In Polity Press.
- [Cu14] Cuesta, M.; Witt, A.K.: How Gender Conscious Pedagogy in Higher Education Can Stimulate Actions for Social Justice in Society. *Social Inclusion*, S. 12-23, 2014.
- [Da14] Dasgupta, N.; Stout, J.G.: Girls and Women in Science, Technology, Engineering, and Mathematics. In *Policy Insights from the Behavioral and Brain Sciences*, S. 21-29, 2014.
- [De15] Denner, J.; Lyon, L.; Werner, L.: Does Gender Matter? Proc. of the 3rd Conference on GenderIT (GenderIT '15), S. 44-48, 2015.
- [El07] El-Nasr, M.S.; Yucel, I.; Zupko, J.; Smit, T.B.: Middle-to-High School Girls as Game Designers - What are the Implications? *Academic Days '07*, 2007.
- [Eu16] European Commission. <http://ec.europa.eu/social/main.jsp?catId=1223>. Stand: 21.03.2020, 2016.
- [Eu19a] European Statistics Eurostat: <https://ec.europa.eu/eurostat/de/web/products-eurostat-news/-/DDN-20190513-1>. Stand: 22.03.2020, 2019a.
- [Eu19b] European Statistics Eurostat <https://ec.europa.eu/eurostat/web/products-eurostat-news/>. Stand: 22.03.2020, 2019a.
- [Fo15] Formanowicz, M.; Cislak, A.; Horvath, L.; Sczesny, S.: Capturing Socially Motivated Linguistic Change. Different effects of gender-fair language on support for social initiatives in Austria and Poland. *Frontiers in Psychology*, 1617, 2015.
- [Ga15] Gabay-Egozi, L.; Shavit, Y.: Gender Differences in Fields of Study: The Role of Significant Others and Rational Choice Motivations. *European Socio*, S. 284-297, 2015.
- [Ga14] Galdi, S.; Cadinu, M.; Tomasetto, C.: The roots of stereotype threat: when automatic associations disrupt girls' math performance. *Child Development*, S. 250-263, 2014.
- [Gi14] Giannakos, M.N.; Jaccheri, L.; Leftheriotis, I.: Happy Girls Engaging with Technology: Assessing Emotions and Engagement Related to Programming Activities: Learning and Collaboration Technologies, S. 398-409, 2014.
- [Go13] GOV.UK <https://www.gov.uk/government/publications/national-curriculum-in-england-computing-programmes-of-study/>. Stand: 21.02.2020, 2013.
- [Ha13] Hamari, J.; Koivisto, J.: Social Motivations To Use Gamification: An Empirical Study Of Gamifying Exercise. *ECIS 2013 Completed Research*. 2013
- [Ha04] Haines, L.: Why are there so few women in games? Digital Games Research Conference, Changing Views: Worlds in Play, S. 1-2, 2004.

- [Hu04] Hunicke, R.; Leblanc, M.; Zubek, R.: MDA: A formal approach to game design and game research, AAI Workshop - Technical Report, 2004.
- [Ka15] Kafai, Y.; Vasudevan, V.: Hi-Lo tech games: crafting, coding and collaboration of augmented board games by high school youth, International Conference on Interaction Design and Children, S. 130-139, 2015.
- [La08] Lambert, L.; Bell, T.; Cortina, T.: Computer Science Unplugged. SIGCSE 2008 Proc. 39th ACM Technical Symposium on Computer Science Education, 553, 2008.
- [La18] Lawlor, J.; Conneely, C.; Oldham, E.; Marshall, K.; Tangney, B.: Bridge21: teamwork, technology and learning. Technology, Pedagogy and Education, S. 211-232, 2018.
- [Ma16] Master, A.; Sapna, C.; Meltzoff, A.N.: Computing Whether She Belongs: Stereotypes Undermine Girls' Interest and Sense of Belonging in Computer Science. Journal of Educational Psychology, S. 424-437, 2016.
- [MA14] Mayring, P.: Qualitative Content Analysis. Theoretical Foundation, Basic Procedures and Software Solution. In basic procedures and software solution, 2014.
- [Mc17] McLean, M., Harlow, D.: Designing Inclusive STEM Activities: A Comparison of Playful Interactive Experiences Across Gender, Proc. 2017 Conference on Interaction Design and Children, S. 567-574, 2017.
- [Me17] Medel, P.; Pournaghshband, V.: Eliminating Gender Bias in Computer Science Education Materials, Proc.2017 Tech Symp. on CS Education, S. 411-416, 2017.
- [Pa15] Paderewski, P.; Arenas, M.G.; Iranzo, R.G.; González, C.; Ortigosa, E.M.: Bringing closer women to engineering: projects and strategies that promote their inclusion, Proc. XVI International Conference on Human Computer Interaction, S. 1-37, 2015.
- [Pa85] Papert, S.: Mindstorms. Children, Computer, and Powerful Ideas. In Basic Books, 1985.
- [Ra20] Raspberry Pi Foundation. <https://www.codeclubworld.org/>. Stand: 04.04.2020.
- [Sp20] Spieler, B.; Oates-Induchová, L.; Slany, W.: Female Teenagers in Computer Science Education. Journal of Women and Minorities in Science and Engineering, *in press*.
- [Sp18a] Spieler, B.; Slany, W.: Female Teenagers and Coding: Create Gender Sensitive and Creative Learning Environments. Proc. of Constructionism 2018, S. 625-636, 2018b.
- [Sp18b] Spieler, B.; Slany, W.: Game Development-Based Learning Experience: Gender Differences in Game Design. Proc. 12th EU Conference on GBL, S. 616-625; 2018c.
- [Sp17] Spieler, B. et al.: Evaluation of Game Templates to support Programming Activities in Schools, Proc. 11th European Conference on Games Based Learning, S. 600-609, 2017.
- [St10] Starruß, I.: Synopsis zum Informatikunterricht in Deutschland. <https://dil.inf.tu-dresden.de/schulinformatik/informatikunterricht-in-deutschland/>. Stand: 02.03.2020.
- [St18] Stoet, G.; Geary, D.C.: The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. Psychological Science, S. 581-593, 2018.

## Programmier-Praktikum für Erstsemester – Erfahrungen aus mehreren Iterationen

Markus Dahm <sup>1</sup>, Jennifer Rose<sup>2</sup>, Marius Köhler<sup>3</sup>

**Abstract:** Die Wirklichkeit in der Programmier-Ausbildung von Erstsemestern hat immer wieder überraschende Erkenntnisse gebracht. Seit 2014 wurden sowohl ein didaktisches Konzept als auch eine unterstützende integrierte Entwicklungsumgebung (5Code) entwickelt, die speziell ProgrammieranfängerInnen unterstützen soll. Das didaktische Konzept vermittelt von Anfang an, wie man in fünf Schritten *vom Problem zum Programm* kommt: *Lesen* → *Verstehen* → *Überlegen* → *Aufschreiben* → *Codieren*. In bisher sechs Iterationen wurden das Konzept und 5Code aufgrund der Evaluation immer wieder angepasst und neu implementiert. Einige Erfahrungen daraus, sowie die daraus gezogenen Schlüsse für die weitere Verbesserung werden in diesem Praxisbeitrag vorgestellt.

### Keywords:

Programmieren, Lernen, Anfänger, IDE, Entwickeln, Hochschule, Sekundarstufe, Software-Entwicklung, Didaktik der Informatik, Kognitive Belastung, Erfahrungen, Medienkompetenz

## 1 Ziel


### 1.1 Vom Problem zum Programm in fünf Schritten

Unser Ziel besteht darin, dass AnfängerInnen sowohl die Vorgehensweise vom Problem zum Programm begreifen, als auch, dass sie diese praktisch anwenden können.

Lernende eignen sich die Syntax und Semantik von Programmiersprachen oft schnell an, wissen dann jedoch oft nicht, wann und wie das theoretisch Gelernte auch praktisch angewendet werden soll [FV10], [Wi14], [Hu07]. Dazu gehört auch, dass sie sowohl die Aufgabe verstehen als auch Mittel und Wege kennen, wie sie zu einer Lösung kommen.

Übliche Entwicklungsumgebungen bieten da kaum Hilfestellung, da diese nur Code behandeln. Bestehende Angebote speziell für ProgrammieranfängerInnen, wie BlueJ, Greenfoot oder Scratch, bieten Unterstützungen nur beim Codieren, decken jedoch Bereiche wie Problemanalyse und Dokumentation von eigenen Ideen und Lösungsansätzen nicht ab [Da15]. Diese sind für den kompletten Ablauf grundlegend wichtig, werden aber vor allem von Anfängern selten wahrgenommen und daher unterschätzt [Hu07], [Lo16].

---

<sup>1</sup> HS Düsseldorf, FB Medien, Münsterstraße 156, 40476 Düsseldorf, [markus.dahm@hs-duesseldorf.de](mailto:markus.dahm@hs-duesseldorf.de),   
<https://orcid.org/0000-0000-0000-0000>

<sup>2</sup> HS Düsseldorf, FB Medien, Münsterstraße 156, 40476 Düsseldorf, [jennifer.rose@hs-duesseldorf.de](mailto:jennifer.rose@hs-duesseldorf.de)

<sup>3</sup> HS Düsseldorf, FB Medien, Münsterstraße 156, 40476 Düsseldorf, [marius.koehler@study.hs-duesseldorf.de](mailto:marius.koehler@study.hs-duesseldorf.de)

Daher vermittelt das von uns entwickelte didaktische Konzept von Anfang an die Vorgehensweise *vom Problem zum Programm* in für Anfänger verständlichen fünf Schritten:

*Lesen* → *Verstehen* → *Überlegen* → *Aufschreiben* → *Codieren*    kurz: *LVÜAC*

LVÜAC betont vor allem die ersten, vorbereitenden Schritte. Codieren wird dabei bewusst als letzter Schritt dieser Kette dargestellt, da es sehr häufig für den einzigen Schritt bei der Software-Entwicklung gehalten wird. Natürlich gehört auch das Testen zur Software-Entwicklung, es steht aber vor allem in den ersten beiden Semestern nicht im Vordergrund.

## 1.2 Reduktion der kognitiven Belastung

Die kognitive Belastung nach der Cognitive Load Theory von Sweller und Chandler [CS96] bei der Lösung von Aufgaben ist für Programmieranfänger sehr hoch [Wi14]: Im kapazitiv sehr begrenzten Arbeitsgedächtnis, das den kognitiven Kontext umfasst, müssen gleichzeitig 1) die Aufgabenstellung, 2) das Verständnis der Aufgabe, 3) die semantischen und 4) syntaktischen Mittel der Programmiersprache gehalten werden. In diesem Kontext soll 5) ein Lösungskonzept entwickelt werden und dann auch noch 6) fehlerfrei und richtig formatiert eingetippt und dabei 7) eine (komplexe) IDE genutzt werden. Die Reduktion dieser kognitiven Belastung ist daher ein wesentlicher Ansatz, um den Lernerfolg von ProgrammieranfängerInnen zu verbessern.

## 2 Iterationen im praktischen Einsatz

Um die kognitive Belastung der Lernenden wirksam zu vermindern und so den Lernerfolg zu erhöhen, wurde die IDE 5Code entwickelt, die über alle fünf Schritte den gesamten Kontext integriert; von der Aufgabenstellung über eigene Notizen bis zur Codierung. Inhalte können dabei über alle Dokumente miteinander verknüpft werden [Da15], [Da16]. Das didaktische Konzept und 5Code wurde ab dem Wintersemester 2014/15 im Praktikum Objektorientierte Programmierung 1 im Studiengang B.Sc. Medieninformatik am Fachbereich Medien der Hochschule Düsseldorf mit 86 bis 104 Studierenden eingesetzt.

In jedem Semester wurden zweimal evaluiert (Mitte und Ende des Semesters mittels eines Online-Fragebogens) sowie das Verhalten im Praktikum beobachtet und dokumentiert. Als Konsequenz wurden Konzepte und Umsetzungen angepasst, verworfen oder neu hinzugefügt. Die Vorgehensweise entspricht damit einem Design-Based Research [An12].

Zu Beginn (2014 bis 2016) wurde eine sehr strikte Struktur der Vorgehensweise vom Problem zum Programm vorgegeben: Die Aufgabe steht in einer eigenen Spalte links, dort kann Text markiert werden, der als Notiz in einer eigenen Spalte daneben organisiert werden kann. Aus jeder Notiz wird automatisch in der Code-Spalte ein Kommentar, der mit eigenem Code ergänzt werden soll. Diese Struktur bildet die Phasen LVÜAC so sehr klar in unterschiedliche, streng voneinander getrennte Artefakte ab (Abb.1).



## 2 Das kleine Einmaleins

a) Geben Sie das **Einmaleins von 1 bis n** auf die Konsole in Form einer **Tabelle** aus. Die Obergrenze n soll dabei variabel sein (d.h. Sie definieren dafür eine Variable).  
Verwenden Sie dafür verschachtelte for-Schleifen. Dabei definieren **anstelle** Sie die Schleifenvariablen im **ersten** vorgegebenen Bereich im

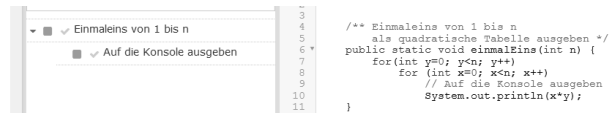


Abb. 1: 5Code – 2014-16 – Aufgabe mit Markierungen, Eigene Notizen, Code mit Kommentar

Das Feedback der Studierenden führte in den Jahren 2017 bis 2018 zur Version *5Code.Local* [Ki17]: Um die eigenen Überlegungen mit dem Code stärker zu integrieren, haben wir *Flow-Programming* als eine aktualisierte Version des „Literate-Programming“ [Kn84] entwickelt [Ro20]. Die Artefakte aller Phasen (Notizen und Code) werden im Fluss (*Flow*), d.h. in der Reihenfolge ihrer Bearbeitung in einem einzigen Dokument integriert. Lesen, Schreiben, Codieren und Orientieren geschieht so im gleichen Kontext, die kognitive Belastung wird so verringert. Außerdem können Code-Blöcke ohne den in Java nötigen Klassen-Code sofort ausgeführt werden. Das vereinfacht den Umgang mit einfachen prozeduralen Programmstrukturen und vereinfacht das Explorieren und Testen (Abb.2).

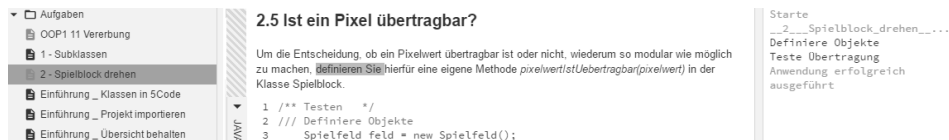


Abb. 2: 5Code.Local - 2017-18: Projekt-Navigation, Aufgabe mit Test-Code-Block, Klassen-Code

Das Konzept *Flow-Programming* wurde im Praktikum positiv evaluiert, allerdings zeigte sich, dass dieses neuartige Konzept einer besonderen Unterstützung bei der Navigation über mehrere Dokumente (Aufgaben und Lösungen) hinweg bedarf (s. 4.2). Daher wurde 2019 die Version *5CodeTNG* neu konzipiert und implementiert [Kö19], [Ro20] (Abb.3).



Abb. 3: 5CodeTNG - 2019: Code-Explorer, Aufgabe mit Test-Code-Block, Klassen-Code

Die Navigation sowohl über Dokumente als auch über den Code wurde vereinfacht und verbessert. Aufgaben und Lösungen werden in zwei Dokumenten angelegt, die beim Scrollen gegenseitig synchronisiert werden („SynchScroll“). Das vereinfacht insbesondere die Arbeit bei Objekt-orientierten Themen stark.

## 3 Erfolgreiche Konzepte

Die folgenden innovativen Konzepte haben sich im Verlaufe der Iterationen bewährt, wie die Evaluationen zeigen, sie wurden jeweils neu entwickelt oder stetig weiter verbessert.

Das Konzept *LVÜAC - Lesen → Verstehen → Überlegen → Aufschreiben → Codieren* hat sich als sinnvoll und nutzbringend über alle hier beschriebenen Iterationen erwiesen. Zunächst wird das Aufschreiben vor dem „eigentlichen“ Codieren als lästig empfunden, gegen Ende des Semesters haben die meisten Studierenden aber den Nutzen erkannt. Vor allem bei Fragen nach Hilfe im Praktikum wurde ihnen deutlich, dass die Beantwortung vereinfacht wird, wenn sie ihr Verständnis der Aufgabe und ihre eigenen Überlegungen kurz aufgeschrieben hatten und nicht nur falschen Code vorweisen konnten.

Die *Integration aller Dokumente über alle Phasen*, von den Aufgaben über eigene Notizen bis hin zu eigenem Code wurde als sehr hilfreich wahrgenommen und auch gerne genutzt.

*Flow-Programming* wurde schnell verstanden und gerne genutzt. Dass Lösungs-Code direkt im Aufgabentext steht, ist bei kleinen und einfachen Aufgaben von großem Vorteil: Der Zusammenhang ist sofort ersichtlich, es muss nichts gesucht werden. Zusätzlich kann der Code reduziert werden auf die zu übenden Bestandteile (z.B. eine Berechnung, Schleife oder Verzweigung). Die Konzentration auf die wesentlichen Code-Bestandteile bietet eine große kognitive Entlastung, die auch in allen Evaluationen immer positiv bewertet wurde. Weiterhin kann jeder Code-Block im Aufgabentext auch direkt ausgeführt werden, wodurch Rückmeldungen und ggf. Hinweise zu Fehlern sofort zur Verfügung stehen.

## 4 Lessons Learned

Nicht alle Erfahrungen sind positiv, nicht alle sind überraschend, sie führten aber immer dazu, Konzepte und Umsetzungen nach jeder Iteration zu überdenken.

### 4.1 Flexibilität erfordert zusätzliche Orientierung

Das prinzipiell einfache und positive Konzept des Flow-Programming wurde erst mit der Einführung der Verwendung von Klassen schwierig, da dann Lösungs-Code sowohl im Code-Block im Aufgabentext, als auch im regulären Java-File stehen kann. Positiv daran ist natürlich, dass man Code immer im aktuell gelesenen File schreiben kann und man nicht ständig zwischen Aufgaben-File und Lösungs-Code-File hin und her wechseln muss. Es entstand aber mitunter Verwirrung, wo denn nun der „eigentliche“ Code stand, da es normalerweise genau ein File für den Lösungs-Code gibt.

Flexibilität muss also noch ausführlicher erläutert werden. Die Studierenden müssen in die Lage versetzt werden, sich ein mentales Modell von der Funktionsweise zu machen [Lo16]. Ggf. ist auch der Rückbau von „gut gemeinten“ Features zu erwägen. In jedem Fall müssen die Orientierung und der Zusammenhang immer gewährleistet werden.

Der später im Studium notwendige Übergang von der Nutzung der speziellen IDE 5Code zu einer üblichen IDE muss ebenfalls beachtet werden. Andernfalls gehen die im besten Fall gut verinnerlichteten Schritte LVÜAC in den rein Code-orientierten IDEs wieder unter.

#### **4.2 AnfängerInnen sind anders anders**

Eine wichtige Beobachtung ist, dass Studierende immer auf ganz eigene Art und Weise Aufgaben bearbeiten. Vor allem die zentrale Zielgruppe der ProgrammieranfängerInnen hat nicht immer die eigens für sie erdachten Hilfsmittel genutzt. So wurden beispielsweise der Code-Explorer (siehe Abb. 3) und selbst die Text-Suche selten genutzt. Das kann mit der Unerfahrenheit und Unsicherheit der Anfänger erklärt werden, die bei (der ja eigentlich zu vermeidenden) kognitiver Überforderung auf bekannte Aktivitäten zurückfallen.

Features müssen daher auf geeignete Art und Weise AnfängerInnen nahegebracht werden. Vielleicht sollten neue Features und Konzepte häufiger in die inhaltliche Aufgabenstellung mit einbezogen werden, um ihre entlastende Funktion laufend sichtbar zu machen.

#### **4.3 Nicht-AnfängerInnen wollen ernst genommen werden**

Die Gruppe unserer Studierenden im ersten Semester ist recht heterogen. Die Spanne reicht von völligen AnfängerInnen (ca. 45%), die noch nie irgendetwas programmiert haben, bis hin zu denen, die bereits aus der Schule, Ausbildung oder Studium viel Erfahrungen mit „Programmieren“ haben, d.h. auch mit üblichen IDEs (ca. 35%).

Wer sich bereits für kompetent hält, akzeptiert die didaktischen Konzepte weniger. Studierende mit Vorwissen und Erfahrungen mit anderen IDEs müssen daher passend „abgeholt“ werden. Anfängern und Erfahreneren gleichermaßen gerecht zu werden, ist eine Herausforderung, die wir in der weiteren Entwicklung angehen wollen, z.B. durch Auto-Adaption oder Adaptierbarkeit der IDE oder auch Gamification-Ansätze („Freischalten“).

#### **4.4 Medienkompetenz und Selbständigkeit scheinen eher abzunehmen**

Die Beobachtungen der BetreuerInnen im Praktikum deuten übereinstimmend darauf hin, dass über die letzten zehn Jahre hinweg Studierende immer leichter überfordert werden, sowohl vom Inhalt als auch vom technischen Umgang mit einem Programm. Die Selbständigkeit bei der Lösungsfindung scheint auch abzunehmen. Immer mehr scheinen kaum Erfahrung in der Benutzung von Software, des Betriebssystems und der Verwaltung von Dateien zu haben, die über das Browsen des WWW und etwas Schreiben hinausgeht. Die Funktionalität wurde insgesamt in 5Code schon gegenüber professionellen IDEs wie VS Code, eclipse, oder IntelliJ, stark reduziert – aber vielleicht noch nicht genügend.

Inhalte, die Art und Weise der Vermittlung und auch die verwendeten Tools müssen daher weiter an Fähigkeiten der Studierenden angepasst werden sollten. Auch sollten nicht zu viele grundlegende Fertigkeiten im Umgang mit Software vorausgesetzt werden.

Funktionalitäten sollten für ProgrammieranfängerInnen schrittweise eingeführt werden sollten, um die Komplexität der IDE und die kognitive Belastung minimal zu halten. Features sollten nur schrittweise freigeschaltet, d.h. überhaupt zugänglich gemacht, werden.

## 5 Ausblick

Aufbauend auf den geschilderten Erfahrungen wollen wir in den folgenden Semestern einiges weiter verbessern, z.B.: detailliertere Rückmeldungen bei der Fehlerbehebung und zum Code-Aufbau, Visualisierung des Ablaufs eines Programms und seiner Effekte, Visualisierung des Lernfortschritts, verbunden mit Elementen der Gamification.

Die AutorInnen bedanken sich für die Förderung als Innovation in der Digitalen Hochschullehre in der Gemeinsamen Programmlinie von MIWF/NRW und Stifterverband.

## Literaturverzeichnis

- [An12] Anderson, T., Shattuck, J.: Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), S. 16-25, 2012.
- [CS96] Chandler, P., Sweller, J.: Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*. 10, S. 151-170, 1996.
- [Da05] Dahm, M., *Grundlagen der Mensch-Computer-Interaktion*, Pearson, München, 2005.
- [Da15] Dahm, M., Barnjak, F., Heilemann M., *5Code - Eine integrierte Entwicklungsumgebung für Programmieranfänger*, In: Hans Pongratz, Reinhard Keil (Hrsg.): *Die 13. E-Learning Fachtagung Informatik (Delfi)*, LNI, Seite 119-130, Gesellschaft für Informatik, 2015.
- [Da16] Dahm, M., Barnjak, F., Heilemann M., *5Code – An Integrated Programming Environment for Beginners*. *i-com J of Interactive Media*, Ausgabe 2-2016, DeGruyter, 2016.
- [FV10] Ford, M, Veneme, S, *Assessing the Success of an Introductory Programming Course*, *Journal of Information Technology Education*, Volume 9, 2010.
- [Hu07] Hubwieser, P., *Didaktik der Informatik*, Springer, Berlin, 2007.
- [Ki17] Kirchhof, D., *5Code.Local- Eine integrierte Entwicklungsumgebung für Programmieranfänger*, Hochschule Düsseldorf, FB Medien, 2017.
- [Kn84] Knuth, D. E., *Literate Programming*. Leland Stanford Junior University, United States: Center for the Study of Language and Information, 1984.
- [Kö19] Köhler, M., *Optimierung der Navigation der Entwicklungsumgebung 5Code*, Hochschule Düsseldorf, FB Medien, 2019.
- [Lo16] Loksa, D. Ko, A. J., Jernigan, W., Oleson, A., Mendez, C. J., and Burnett., M. M., *Programming, Problem Solving, and Self-Awareness: Effects of Explicit Guidance*. In *Proceedings of the 2016 CHI ACM.*, New York, NY, USA, S. 1449–1461, 2016.
- [Ro20] Rose, J.: *Optimierung des Workflows der Programmierumgebung 5Code für Programmieranfänger*, Hochschule Düsseldorf, FB Medien, 2020.
- [Wi14] Williams, J.S: *A Computer Learning Environment for Novice Java Programmers That Supports Cognitive Load Reducing Adaptations and Dynamic Visualizations of Computer Memory*, Dissertation, Paper 574, Univ. of Wisconsin, Milwaukee, 2014.

## Automatic Feedback for Open Writing Tasks: Is this text appropriate for this lecture?

Sylvio Rüdian<sup>1</sup>, Joachim Quandt<sup>2</sup>, Kathrin Hahn<sup>2</sup> and Niels Pinkwart<sup>3</sup>

**Abstract:** Giving feedback for open writing tasks in online language learning courses is time-consuming and expensive, as it requires manpower. Existing tools can support tutors in various ways, e.g. by finding mistakes. However, whether a submission is appropriate to what was taught in the course section still has to be rated by experienced tutors. In this paper, we explore what kind of submission meta-data from texts of an online course can be extracted and used to predict tutor ratings. Our approach is generalizable, scalable and works with every online language course where the language is supported by the tools that we use. We applied a threshold-based approach and trained a neural network to compare the results. Both methods achieve an accuracy of 70% in 10-fold cross-validation. This approach also identifies “fake” submissions from automatic translators to enable more fine-granular feedback. It does not replace tutors, but instead provides them with a rating based on objective metrics and other submissions. This helps to standardize ratings on a scale, which could otherwise vary due to subjective evaluations.

**Keywords:** Feedback; online courses; language learning.

### 1 Introduction

In second language learning giving feedback to students is of high importance to enable them to reach the desired learning outcome. While closed tasks, e.g. filling blanks, can be easily checked for correctness due to existing sample solutions, open writing tasks often need to be assessed by tutors to provide appropriate feedback for students.

Research shows that language levels can be derived from written texts; the results show the language level (A1, A2, ...) of the submitted student’s text [K184]. When observing online language learning courses of individual language levels, all students generally learn the same words and grammatical structures as defined in the “Common European Framework of Reference for Languages” [Co01]. However, course book editors have a degree of freedom to decide which topic and which grammar will be taught within each course at any individual level. This results in a diversity of online courses with different structures. However, simply obtaining the information that a student’s written text meets the requirements of a particular level is not enough to give automatic feedback, apart from in the

---

<sup>1</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Weizenbaum Institute for the Networked Society, ruediasy@informatik.hu-berlin.de, <https://orcid.org/0000-0003-3943-4802>

<sup>2</sup> Goethe-Institut e.V., Oskar-von-Miller-Ring 18, 80333 Munich

<sup>3</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, pinkwart@hu-berlin.de

extreme scenario that a student's text does not fit into the current language level at all or when students are not aware, especially at the beginning of the course, that they are required to write texts that reach the language level of their current course.

Many automatic essay scoring systems only take samples of texts with existing ratings into account, but miss the relation to an online course where users have been taught other words, expressions, and grammar; i.e., learners may have extended their knowledge that should be applied in any particular practical open writing task. Thus, we aim to bridge the gap and focus on features that are related to taught contents in combination with already known submissions to get a more in-depth insight to predict teachers' ratings of student texts. In this paper, we follow the research question: which generalizable method could be used to predict tutor ratings of open writing tasks in online courses by combining the meta data collected from such tasks with the actual language content of any particular course section.

## 2 Related Work

Alikaniotis et al. [AYR16] introduced an approach of automatic text scoring in language learning using a long short-term memory (LSTM) neural network. Rating typically depends on a set of textual features such as grammar, vocabulary, style, relevance, and complexity of sentences. According to the authors, regression and ranking are still the state of the art approaches to obtain automatic ratings that are indistinguishable from human ratings. Using their architecture of an LSTM, they achieve comparable accuracies with fine-tuned models. With a feature importance analysis, they state that they were able to find features that have an impact on predictions, but there is no causality to suggest that these features cause different ratings.

Klebanov et al. [KF13] introduced new features using the co-existence of words in a text for automatic essay scoring; these features show a relationship between the quality of writing and word association profiles. Automatic essay scoring was also investigated by McNamara [Mc15] using a hierarchical classification and predict scores based on competency levels. They state that a hierarchical approach has the advantage of having the ability to inform by providing general formative feedback. Truscott [Tr96] shows that it is not necessary to give feedback for every single mistake in language learning as this kind of correction is unnecessary, ineffective and expensive. Teachers have to decide how fine-granular a correction should be; a detailed correction, for example, is always more time-consuming and time itself is a limiting factor. Having a method for giving a less-detailed, more general automatic feedback, could help to reduce tutor workload and also makes it possible to give immediate general feedback, without the need to wait for the teacher's rating. The usefulness of getting detailed or less-detailed feedback is a much debated topic amongst researchers, with little agreement being reached. Beuningen et al. [BJK08] notice that direct corrective feedback has a long-term effect, but short-term effects were observed

for both direct and indirect corrective feedback. There is no need to discuss whether feedback is necessary, but providing good feedback in online courses on a large scale requires either a considerable number of teachers or good technology-driven approaches that enable a feedback almost comparable to that of a tutor.

In open writing tasks, computer-generated submissions are a problem where learners send a text that they claim to have written themselves. Lavoie et al. [LK10] show that texts can be classified relying on keyword-based features. This is an important step to identify “fake” submissions, instead of rewarding a user for a good text. The authors recognize that there is still considerable room for improvement in detecting fakes.

In this paper, we propose a methodology to give language learning students feedback for the appropriateness of written texts for every section of an online course to learn German as a foreign language. This approach is generalizable and can be applied for every language learning course for languages that are supported by our used libraries. Additionally, it can detect fake submissions.

### **3 Methodology**

Our prediction combines different layers of language learning. We explore the use of grammar and wording separately, and also combine them to find the most practicable accuracy.

#### **3.1 Grammar & Mistakes**

The first focus is on the appropriateness of the applied grammatical structures that students were taught during the lecture. A “trivial”, but very comprehensive and expensive method would be to define, which grammatical structures the learner should be able to use, for each section of a language learning course. As this method requires language experts, we do not want to limit ourselves to their availability. Instead, we propose a greedy algorithm by observing the receptive parts of the online course; this includes texts, where learners have to solve tasks. We follow the assumption that learners know how to use at least some of the grammatical structures that have been taught in the particular section involved.

To determine whether grammatical structures were used in open writing tasks, we first extract all receptive texts of a particular section. We transfer all texts to their Part-of-Speech (POS) Tags representation [He99]. Each word will be transferred to the corresponding part of speech (e.g. NN for a noun). This allows us to have access to a generalized version, thus obtaining a list of word- and punctuation classes. Subsequently, we define n-grams and set n to three, named trigrams. These trigrams consist of three connected sequential POS tags. Consecutive parts of speech tags represent possible grammatical structures that the user was confronted with. The same is done with previous submissions of users that have previously been marked by tutors as appropriate for the section.

The next step is to find overlapping POS trigrams of the section and the user’s submission. This set contains grammatical structures that are part of the online course section itself and those that the learner used within the submitted text. For every existent acceptable submission of our observed section, we are able to obtain a set of trigrams. If we count the number of overlapping trigrams as values, we can put them into a sorted list, that represents the range of acceptance according to the grammatical structure. By finding the overlap with the POS trigrams of the section and a new user’s submission, it is possible to see whether the resulting length of the intersection will be in our learned range.

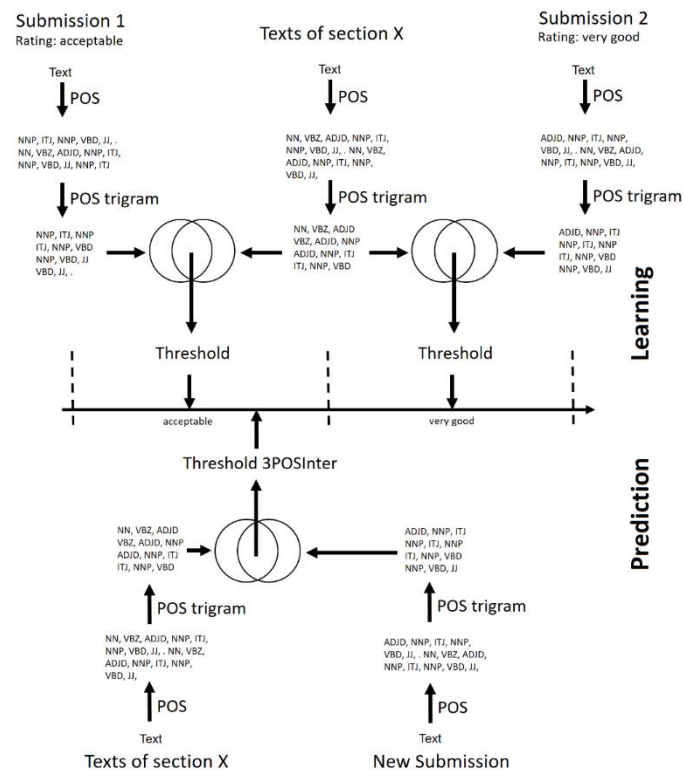


Fig. 1: Methodology for using thresholds of POS trigrams intersections.

This learned range can help to understand how good a submission is. We tag this value with “3POSInter”. Fig. 1 visualizes the learning and prediction processes in general. The didactic approach follows the standard criteria for correction of written texts for German as a foreign language<sup>4</sup>:

<sup>4</sup> [https://www.goethe.de/lrn/pro/gc1/C1\\_PruferTraining\\_08.pdf](https://www.goethe.de/lrn/pro/gc1/C1_PruferTraining_08.pdf)



1. All parts of the task have been answered.
2. Content is presented in a coherent and well-structured way.
3. Vocabulary used is matching the current level of the course and according to the level of the Current European Framework of Reference for Languages [Co17].
4. The text is correct regarding morphology, syntax, orthography and punctuation.

While criteria 1 and 2 are implying content analysis and semantic understanding of the text, criteria 3 and 4 are orientated on formal aspects of language. As criteria 3 and 4 are easier to analyze with the technology in use, focus will be placed on these two criteria. We divided the acceptance range into three classes and added smileys for each class for visualization of general feedback: ☹=acceptable, 😊=good, 😄=very good. If the POS trigram overlap is smaller than the known range, the tool gives ☹=bad as the feedback. The “very good” class is the range starting at the third of the range vector and is open-ended.

Practical usage has shown that learners who write more, are advantaged by this approach. Using grammar that the user should not be aware of at this stage could represent a fake submission, possibly achieved by using a translation tool, for example. To filter out these submissions, we used a superset of the sections’ and submissions’ trigrams; this superset contains all trigrams that occur in the submission of the user that are not part of the online course. We computed the length of this superset for all known submissions of this section to see whether the range is appropriate or not. This is labeled “3POSSuper”. We use the end of this range as a fixed border. If the supersets length of a new submission is greater than this border, we have a solution that contains too many mistakes. This means either the learner uses grammatical structures that were not taught during the lecture or uses inadequate grammar that results in trigrams that do not exist.

To distinguish between “fake” submissions and an unreasonably high number of mistakes, we added the last step. We use all texts at every level of the online course to obtain the POS trigrams. We follow the assumption, that, if the trigram generated is not part of the online course (ranging from A1 to C1), it is seen as non-existent and should not be used. Computing the superset of the trigrams of the overall online course and the learner’s submission gives the amount of possible non-existent grammar structures, which are then counted as mistakes. Fake submissions have very few of these mistakes; there are usually fewer trigrams that are non-existent in the overall online course compared to other submissions. Thus, the submission can be flagged as out of scope and should be revised by a tutor. Having numerous mistakes in a submission can be flagged as ☹=bad and should be revised by the learner.

The following paragraphs formalize our approach. Let  $\{T_{i,lec}\}$  be the set of POS trigrams of the section  $lec$  which contains  $i$  texts  $T_i$ . Let  $\{T_{j,lec}\}$  be the set of POS trigrams of the user’s submission  $j$  for the specific open writing task in the section  $lec$  of the online course. The next step creates vector  $\vec{V}$ , ordered by value, that includes all lengths of the

POS trigrams overlaps between the online course section and each existing user submission:

$$\forall T_i \in lec: v_i = |\{T_{i,lec}\} \cap \{T_{j,lec}\}|$$

$$\vec{V} = sort[v_o, \dots, v_i]$$

Vector  $\vec{V}$  now contains acceptable lengths of intersections, ranging from low performing submissions to very good texts. To have a better orientation, we divide this vector into three parts of equal length: acceptable, medium, very good. Of these parts, we use the middle value and define our lower and upper thresholds: F = acceptable, L = good/very good:

$$F = \vec{V} \left[ \text{int} \left( \frac{\text{len}(\vec{V})}{6} \right) \right], L = \vec{V} \left[ \text{int} \left( \frac{5}{6} \text{len}(\vec{V}) \right) \right]$$

By calculating the length of the POS trigrams overlap with a new submission and the section, we get a metric to compare the users' submissions among each other; this metric can also be used to compare future submissions.

### 3.2 Wording

Next, we used stem words that occur for the first time in the section of the online course and count the length of the intersection with the stem words of the submissions. This is labeled "WORDS". We limit used words to stem words as declined words could be misspelled by users or not recognized by the system, and grammar usage is already covered by the previous metric. In a practical setting, focusing on stem words will result in more hits, whereas using declined words will often result in zero hits, which would defeat the purpose.

To focus on newly learned words of a particular course section only, we collected all stem words of all sections. If a word was already part of a previous section, it was removed from the current one. This resulted in a list, for every section of the course, only containing words that the user was confronted with for the first time in the particular section. In contrast to our approach for grammar, where we used receptive texts only, we use all texts of all sections, independently of them being part of a task or a reading/listening text. It could be observed that people mainly focus on vocabulary learning and submissions with a good rating use learned words of the section, which the learner had not necessarily encountered previously.

Subsequently, we apply different methods to predict the tutor's rating of users' submissions. We searched for correlations and an optimal regression, we use the thresholds as defined above and observe a neural network to compare the accuracy.

## 4 Results & Evaluation

An experienced tutor rated 400 users' submissions of a specific section of the online course "Deutsch Online"<sup>5</sup>. Texts could be labeled with the following ratings: 1 = very good, 2 = good/acceptable, 3 = not acceptable, 4 = too bad/out of scope/fake.

Figure 2 shows all ratings concerning the features we observed. Each dot represents a single user text according to the observed feature. The dataset is as expected. The more POS trigrams that exist in the submission that also occur in the course, the better the rating (Fig. 2 a). Not acceptable submissions (rating 3) use fewer POS trigrams that exist in the course, but submissions that are out of scope contain more trigrams that are not part of the course trigrams (Fig. 2 b). Submissions get bad ratings the more mistakes that are found (Fig. 2 c). Texts that are out of scope (e.g. because of being a fake submission) have fewer mistakes on average, as expected.

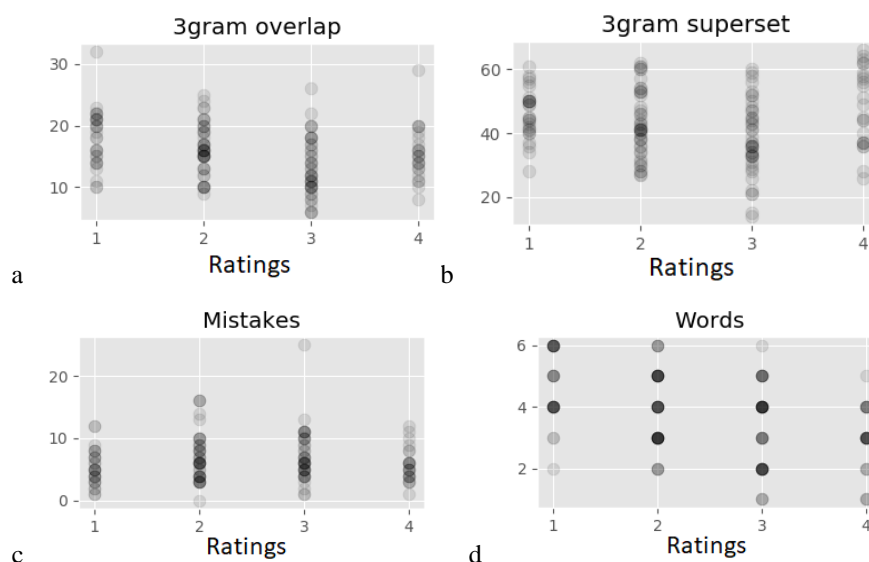


Fig. 2: a) POS trigrams intersection of users' submissions and course content. b) The number of POS trigrams that occur in the users' submissions, but which are not part of the course content. c) The number of found mistakes in the users' submissions. d) The intersection of stem words in the users' submissions, and course content.

The more stem words, which are included in the online course and used by the learner, the better the rating (Fig. 2 d). This general observation shows that there are some remarkable patterns concerning our metrics and ratings. To understand, whether a linear relationship between our meta-data and the rating exists, we used the Pearson Coefficient:

<sup>5</sup> <https://www.goethe.de>

$$r([3POSInter; WORDS; 3POSSuper; MISTAKES]) = [-0.26; 0.45; -0.03; 0.05]$$

This rough analysis shows that our observed superset has almost no linear correlation. Mistakes in submissions have a low impact on the rating. It makes sense that the superset alone is no linear indicator for a rating and it should only be used to determine whether a submission is out of scope. A detailed rating solely using this metric is impossible. The feature “WORDS” has a moderate correlation only and the correlation of 3POSInter is weak. Mistakes also have a very small linear influence on the rating. To see whether other non-linear dependencies are existing, we used a regression approach to predict all ratings. The resulting RMSE shows that this methodology cannot be used for good predictions (standard deviation = 1.003):

$$RMSE([3POSInter; 3POSSuper; MISTAKES; WORDS]) = [1.04; 1.00; 1.01; 1.09]$$

The absolute error is in the range of 0.87 and 0.92, which shows that on average the predicted rating lies in the right direction. Furthermore, combinations of all our features for a multi regression were tested but resulted in  $RMSE = 1.14$ , which is an even higher error rate than for single features.

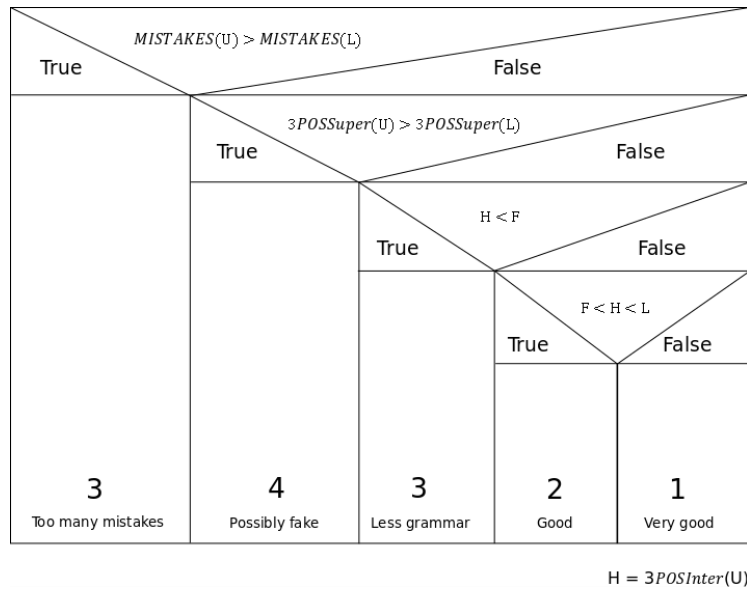


Fig. 3: Hierarchical decision tree for an algorithmic rating of users' submissions using different thresholds.

We can find thresholds according to our metrics and the existing user submissions. According to our findings, all thresholds combined result in a decision tree to rate grammar. It can be found in Fig. 3, with  $U$  as the user's submission,  $L$  as the previously learned

parameters and  $H$  as the intersection of the user submissions' POS tags and the POS tags of the lecture.

Testing this approach in cross-validation shows that 70% of all ratings can be predicted correctly, by using the hierarchical decision tree. In this paper, we do not differentiate between the granularity of good ratings, as the main concern is to find a way to detect acceptable and not acceptable submissions. More granular observations are part of further studies. Using the mean value of existing stem words in the existing user submissions related to each rating also results in fixed thresholds. It cannot be varied much as is visible in Fig. 2 d): user submissions contain up to 6 words that they have previously learned. Thus, more granularity is not possible. Merging rating 1 and 2 to be an acceptable solution and 3 and 4 to be an unacceptable solution results in a more robust rating. In our case:

$$R_{1,2} = \emptyset(\text{Rating1} \cup \text{Rating2}) = 4.2; R_{3,4} = \emptyset(\text{Rating3} \cup \text{Rating4}) = 3.2$$

Using  $R_{1,2}$  as a threshold, 63.2% of the ratings could be predicted. Using  $R_{3,4}$  as threshold results in a prediction accuracy of 63.9%. Each value lower or higher of  $R_{3,4}$  decreases the accuracy. Thus it is recommended to use  $R_{3,4}$  as a threshold.

Having fixed thresholds by looking at our features separately, shows that we reached a limit in prediction concerning our observed features using threshold-based methodologies. To overcome this problem, we use a neural network that gathers all features together to find any possible existing pattern amongst them. As this is a black-box approach, we can only see how good the prediction works, without receiving any insights into the internal patterns. Our neural network has a two-layer architecture with “hard sigmoid” as activation functions and “Nadam” as the optimizer. All available features are used for training (*3POSInter*, *3POSSuper*, *WORDS*, *MISTAKES*). Our data was balanced to avoid overfitting problems. We used Gridsearch [Sc17] to find the optimal hyper-parameters combination. With 1500 epochs and a batch size of 50 we achieved an accuracy of 69.92% in 10-fold cross-validation. Limiting our training data to the feature “*WORDS*” only, the neural network achieves an accuracy of 63.85% in 10-fold cross-validation. Using *3POSInter*, *3POSSuper*, *MISTAKES* as done within the previous approach, we achieve 63.46%. If we use these features separately for training, the achieved accuracy is lower (*3POSInter* : 61.54%, *3POSSuper*: 55.38%, *MISTAKES*: 46.15%). By combining these features we achieve the best accuracy; this is comparable to the result when using our trained thresholds. To conclude our results, we see that the best achievable accuracy is 70% with both of our used methodologies. The accuracy of using words only is nearly the same for both methods (63.85% and 63.9%); the difference is just a matter of fine-tuning.

## 5 Discussion

Neumeyer et al. [Ne00] measured the human-to-human correlation in rating spoken sentences by teachers. Raters had an inter-rater reliability among each other between 0.61 and

0.72 at the sentence level. Compared to our result, machine-to-human reliability of the same ratings achieved 0.7, which is acceptable for practical applications.

Both of our methodologies show very similar results when predicting tutor ratings for open writing tasks. The threshold approach has the advantage that explainable ratings can be created, which is impossible when using the neural network as a black-box. As the results are very similar, it is recommended to use the threshold approach, allowing for explainable ratings, for further applications. The application of our meta-data extraction requires a separation of the course's texts into receptive parts and new content. This is an important step to achieve meta-data for more granularity but it requires manpower.

Alternatively, to find possible mistakes due to the non-existence of POS trigrams in the overall course, a third-party tool can be used to find mistakes, the quantities found can be used as a feature. This may be necessary if the whole course does not teach and use all existing grammatical constructs of a language. Furthermore, such a tool could also find spelling mistakes that could help in the prediction of ratings.

Even though teacher training and examiner certification have the objective to give guidelines on the correction of written texts, the current definitions, to a certain extent, leave room for interpretation for the correcting teacher. There is no hard limit that can be defined between the different grades, as it is up to the teacher to decide whether a text is in general understandable or whether only parts of it are. Our approach of automatic feedback on written tasks can support teachers if they are undecided as to which of two grades is the best one.

Improvements can be understood in two directions. We expect that teachers will be able to provide more standardized feedback based on the automatic evaluation of learners' texts and they will also be able to focus more on the content itself and less on the correction of formal problems and errors to support the language learning process. In addition, the time spent on the correction of learner texts will be reduced. Having tutoring as the main factor on the cost side of online courses, this could make online courses more financially accessible to learners. Currently, these courses can be found in the premium sector of the language learning market. It is necessary to provide additional information as part of the given feedback to make the understanding of the prediction easier. As in many parts of this proposal, it is important to state that technical analyses and didactic approaches need to be linked to each other.

## **6 Future Work**

We have seen that we are unable to achieve more than 70% accuracy with our used meta-data. Further research could focus on collecting more meta-data to be used as features, but the result may still vary depending on the tutor. As the overall aim is to improve the accuracy of predicted ratings, we would like to explore more features that can be derived from

user submissions. One idea is to get insights into semantics. Our proposed solution only takes grammar and used words into account. Words themselves determine the context, but whether a written text is appropriate with regard to the context has not yet been examined. Taking an algorithmic approach to understand in which relation words are used could be helpful to distinguish between good and unacceptable submissions. Another extension is the combination of words in general, not just searching for new words that only co-occur in the online course. Looking at bi- or trigrams of words gives a more in-depth overview of the context.

Future scenarios beyond the practical use for the correction of learner texts in language courses could involve placement tests. They detect the current language level of the learner and are normally used at the beginning of a language course or to certify a level for job applications. Available automatic tests for German as a foreign language are limited to the receptive skills of reading and writing. The productive skills of writing and speaking are either assessed by a tutor, incurring additional costs, or are simply left out. This can lead to an inaccurate evaluation of the language level which could result in the need for a class change at the beginning of a new course. More precise placement including an automatic evaluation of the learner's level would improve the learning experience, reduce frustration due to being placed at the wrong level and would give more precise information on language qualification for personalized training programs.

The application could cluster typical errors and problem areas, conduct the learner to adaptive training of these errors and generally guide the learning process according to the student's need. Often automatic translations are used and in many cases the learner is unable to understand or evaluate whether the translations are meaningful in the larger context. Moreover, translations may only be helpful to finish off a task but they do not normally support the real learning process. At a more advanced stage, after having proven high accuracy in placement tests, automatic correction could also support official language certification. Exemplarily, the writing section is corrected by two grading persons and in the case of a large discrepancy, a third corrector will review and take the final decision. A trained automatic correction tool could be used to replace the first correction stage and could subsequently be compared to a manual correction, allowing students to receive their final exam marks within a shorter time period. Besides some investigation is required to understand how tutors use automatic rating, whether they guide or manipulate them.

## **7 Conclusion**

In this paper, we explored the prediction of ratings for open writing tasks to support tutors of online language learning courses. The novelty of our approach is the combination of online course contents and existing submissions to derive new features. Our different experiments have shown that we can predict tutor ratings with an accuracy of 70%. The results both using fixed thresholds and a neural network are similar. Thus, for this area of

learning, the more explainable method, i.e. fixed thresholds, should be used. As the predicted result is explainable, generated feedback can be more granular to give recommendations, e.g. to use more newly learned words or to create more complex sentences. Our approach can help tutors in online courses, as the prediction gives tutors advice to enable more consistent ratings among different tutors. Our practical results are promising and observing more features could help to optimize the accuracy in further investigations.

**Acknowledgments:** This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII116 (Weizenbaum-Institute) and the Goethe Institute e.V. The responsibility for the content of this publication remains with the authors.

## Bibliography

- [AYR16] Alikaniotis, D.; Yannakoudakis, H.; Rei, M.: Automatic Text Scoring Using Neural Networks, in *ACL*, 2016.
- [BJK08] Beuningen, C. v.; Jong, N. d.; Kuiken, F.: The Effect of Direct and Indirect Corrective Feedback on L2 Learners' Written Accuracy, in *ITL - International Journal of Applied Linguistics*, Volume 156, Issue 1, 2008.
- [Co01] Council of Europe, A Common European Framework of Reference for Languages: Learning, Teaching, Assessment, in Council for Cultural Co-operation, Strasbourg, Cambridge University Press, 2001.
- [Co17] Council of Europe, The CEFR Levels, 2017, Available: <https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions> [Accessed 31 03 2020]
- [He99] Heeman, P. A.: POS Tags and Decision Trees for Language Modeling, 129-137, 1999.
- [KF13] Klebanov, B. B.; Flor, M.: Word Association Profiles and their Use for Automated Scoring of Essays, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, p. 1148–1158, 2013.
- [Kl84] Klein-Braley, C.: Advance Prediction of Difficulty with C-Tests. 1984
- [LK10] Lavoie, A.; Krishnamoorthy, M.: Algorithmic Detection of Computer Generated Text, 2010.
- [Mc15] McNamara, D. et al.: A hierarchical classification approach to automated essay scoring, in *Assessing Writing* Volume 23, Elsevier Ltd., 2015, pp. 35-59.
- [Ne00] Neumeyer, L. et al.: Automatic scoring of pronunciation quality, in *Speech Communication*, Volume 30, Issues 2–3, Elsevier Science B.V., 2000, pp. 83-93.
- [Sc17] Scikit-learn developers, Tuning the hyper-parameters of an estimator, 2017. [Online]. Available: [http://scikit-learn.org/stable/modules/grid\\_search.html](http://scikit-learn.org/stable/modules/grid_search.html). [Accessed 30 07 2018].
- [Tr96] Truscott, J.: The Case Against Grammar Correction in L2 Writing Classes, in *Language Learning: A Journal of Research in Language Studies*, Volume 46, Issue 2, 1996, pp. 327-369.



## Educational Text Summarizer: Which sentences are worth asking for?

Sylvio Rüdian<sup>1</sup>, Alexander Heuts<sup>2</sup> and Niels Pinkwart<sup>3</sup>

**Abstract:** Many question generation approaches focus on the generation process itself, but they work with single sentences as input only. Although the state of the art of question generation’s results is quite good, it cannot be used practically as the selection which sentences are worth asking for in an educational setting is currently not possible in an automated way. This limits the ability to generate interactive course materials at scale. In this paper, we conduct a study where we compare teachers’ sentence selections of texts with 9 algorithms to find the most appropriate ones concerning reading comprehension. 30 teachers compared the “winner” algorithm, Edmundson with LexRank, which was found to be the optimal algorithm according to previous literature. The result shows that Edmundson outperforms LexRank.

**Keywords:** Question generation; Online courses; Text summarization.

### 1 Introduction

Asking questions is one of the main methods used for testing reading comprehension. The preparation of questions is time-consuming and requires an understanding of used texts. The success of interactive online courses depends on a well-structured presentation of contents as well as asking appropriate questions to engage students and test their knowledge.

On the one hand, automatic question generators are available and perform well, but existing approaches of question generators typically focus on generating questions for single sentences based on templates or other methods. However, in many educational scenarios, there is textual material existing that covers many sentences and not just one. The existing question generation techniques cannot select the sentences that are worth asking for in an educational context. This selection task is of critical importance for generating high-quality questions for educational purposes [CYG19]. Doing this at scale can support the process of creating interactive learning material including questions and quizzes.

On the other hand, there are already existing extractive text summarizers that can summarize texts – a question generator could generate questions for selected sentences of the

---

<sup>1</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Weizenbaum Institute for the Networked Society, ruediasy@informatik.hu-berlin.de, <https://orcid.org/0000-0003-3943-4802>

<sup>2</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Weizenbaum Institute for the Networked Society, alexander.heuts@hu-berlin.de, <https://orcid.org/0000-0002-1755-8970>

<sup>3</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, pinkwart@hu-berlin.de

summary. Yet, the resulting summaries depend on the used algorithms, which are designed by using different approaches. Resulting summaries are thus different. Furthermore, the overall goal of text summarizers is to summarize texts by extracting the main aspects. This approach does not necessarily lead to educationally meaningful sources for question generation. To assist instructors and teachers in creating questions for texts concerning reading comprehension, there is the necessity to use an approach that extracts sentences that are worth asking for first and then generate the questions. If an automated summary algorithm produces results that are similar to a teacher's selection, this can thus help to create educationally meaningful questions at scale.

The research objective of this paper is to explore which already existing extractive text summarizers were preferred in sentence selection to ask questions for reading comprehension concerning different subjects that are taught at school.

## 2 Related Work

Question generation (QG) is classical a sequence-to-sequence problem, with a text or a sentence as input and a question as output. Sometimes, the answer can also be a part of the output. Two approaches emerged in recent years to generate questions: rule-based and methods based on neural networks [De18]. Rule-based methods use templates that could be created manually or be generated. At the generation step, they use the syntactic structure of the language to create questions by re-ordering words. The success of these approaches depends on the existence of multiple templates with diverse structures. All QG approaches are limited to generate questions for single sentences, which still require the selection of sentences that are worth asking for. Using the state of the art, generators create one or more questions for every sentence of a given text, without selecting appropriate ones. Current researchers use human evaluation only to rate generated questions by quality, clarity, usefulness, or grammatical correctness [CYG19, DSC17]. But they limit their evaluations to already selected sentences and mix up the evaluation of selected sentences and the resulting generated questions.

Du et al. [DC17] proposed a hierarchical neural sentence-level sequence tagging-model as a data-driven approach to identify sentences that are question-worthy by using a paragraph as input. For training, they used SQuAD to find sentences that are worth asking for. But SQuAD was designed for "machine comprehension of texts" [Ra16] and not for educational purposes. The main idea of machine comprehension is to ask questions to the machine and the algorithm can find the answer within a given text [Ra16]. It also can be used to train a question generator based on single sentences [HS16]. Concluding to be useful for selecting question-worthy sentences is "not suitable for question generation in the learning context" [Ch18] as it was created by crowdworkers, not by teachers with a pedagogical background and without the aim to create questions for reading comprehension at school. The didactical background for asking questions is missing in the dataset. Assuming that the dataset can be used for real-world educational environments cannot be justified

[Ch18]. “Crowdworkers were tasked with asking [...] questions on the content of that paragraph” [Ra16], but a concept to ask questions with the aim of human reading comprehension (e.g. at school) is missing. Although the approach of the authors is a novelty in research, the results need to be revised by using a new database.

G. Chen et al. [CYG19] used four different datasets (TriviaQA [Jo17], MCTest, RACE, and LearningQ) to examine which existing text summarizers select the sentences which were asked for in the samples. All datasets were designed for different purposes. RACE is mostly used in the setting of language learning. LearningQ contains different educational questions of different subjects with 7k questions created by instructors, containing questions to seek for an in-depth understanding of taught concepts. The remaining two were not created for educational settings. The question generator itself was trained with SQuAD [Ra16]. In their evaluation, they used Grammaticality, Clarity, and Usefulness to evaluate generated questions by humans and found out that LexRank performs best.

The overall problem that datasets were used to train models for purposes that they were not created for can result in recommendations that are practically unfeasible. An evaluation of resulting generated questions depends on the quality of the question generator itself, which limits the aimed evaluation of the selection part. In this paper, the main concern is to focus on the selection of sentences for the sake of human reading comprehension to support teachers and instructors to create or generate questions. We focus on two major research questions: 1) Which text summarizer is most similar to the instructor selections for the task of reading comprehension? 2) Does this text summarizer perform better than LexRank, which is the best according to the literature [CYG19]?

### 3 Methodology

#### 3.1 Study Design

Our study consists of a pre-study, followed by a comparison of our findings with the literature's findings to determine the best state of the art algorithm that selects sentences that are worth asking for in an educational environment. Within the pre-study, we compared selected sentences by teachers and automatic text summarizers. We used 48 texts that have been annotated by 3 teachers each to select sentences that are worth asking for concerning reading comprehension. For this study, teachers were requested by an open call for participation to annotate 8 texts each. For this purpose, German-speaking teachers were sought regardless of their subject or federal state. In sum, 18 teachers annotated all texts, each one three times. Besides, teachers also had to formulate a question for each selected sentence with the corresponding answers so that the task was authentic and they were aware of the reasons for selecting sentences. The texts have an average length of 2687 characters, ranging from 1759 to 3731. They are selections of 12 school subjects (German, geography, history, mathematics, physics, chemistry, biology, computer science, art, music, political education, and theories on sports). For each subject, four specific texts were chosen, that

are usually taught at school. The topics of the individual texts are based on the curricula of secondary school subjects from different federal states in Germany. There was no fixed pool of topics from which each subject was chosen randomly. The selection of the topics was made by the instructors, whereby as many different topics as possible were chosen for each subject. Furthermore, the teachers had no influence on the selection or assignment of the topics and texts they were supposed to annotate, as this has been done randomly. Instead of actual school readings, we used extractions of German Wikipedia articles due to copyrights and publishing reasons. The study was planned to be done online, thus using texts by publishing houses was not possible. Texts have been shortened and re-formulated by instructors of the institute for computer science education at the Humboldt University in Berlin, who have partly completed teacher training and are also educating teachers, with the aim to prepare our texts for teaching at school. Strongly interlaced sentences were split into multiple sentences if it was possible. This step was necessary; otherwise, long sentences could be favored as they might contain more information that is worth asking for than shorter ones.

In the main study, we used 9 extractive text summarizers which had the task to select three sentences, more precisely: TextRank [MT04], Edmundson [Ed69], LexRank [ER04], LSA [De90], SumBasic [NV05], Luhn [Lu58], MMR [GC98], KL [HV09], and Longest Sentences. The algorithms had to do their selections on the same texts that we used in our pre-study. According to the length of our texts, asking three questions on a remember level at Bloom's taxonomy [Bl65] is appropriate. The number of selected questions is independent of the used summarizers. Details of all summarizers can be found in the references. We compared the similarity between human selections and the algorithmic ones. Finally, we ordered the similarity scores by value to find the best performing summarizer, that can be used as educational text summarizer to find question-worthy sentences. The question generation step by using single sentences as an input to create a question is not part of this paper as this has already been investigated a lot in detail [CYG19]. With this pre-study, we can detect algorithms that chose sentences most similar to the teachers' selections.

### 3.2 Algorithmic Evaluation

Different metrics are existing that are broadly used to evaluate generated questions or summarized texts. Metrics like Rouge<sub>L</sub> [Li04] can be used to evaluate summarized texts with a reference. This identifies the longest subset of co-occurring sequences of the original sentences concerning generated ones. It uses n-grams recall for evaluation with gold standard sentences as references. Bleu[1-n] uses n-grams' co-occurrences, too, with an additional penalty for overly short sentences [Pa02]. Originally this metric was developed to evaluate a machine-translated text with a human's translation. These sequence-to-sequence problems have in common that texts and summaries or translations can be completely different among each other. By focusing on the micro-level, which sentences are worth asking for, selected sentences from a text will not be changed or transformed. Thus,

in our case, using n-gram approaches are overwhelmed as they were designed for transformed texts and we use extractions only. Statistical-based metrics like Rouge<sub>L</sub> or Bleu are necessary if the teacher’s selection is not known and the comparison takes place between texts and questions. As we have, due to our study design, the knowledge of selected sentences which are worth asking for, we can use the overlap of selected sentences to compare teachers’ selections with selected items by each used summarizer. Using this overlap we define our similarity score ( $sim_A \in [0,1]$ ), where  $sim_A = 1$  means that selected sentences by teachers and the algorithm  $A$  are identical. Let  $Q$  be the number of texts,  $N_Q$  the number of sentences of text  $Q$ , and  $P$  the number of selected sentences per text. Let  $C_L$  be selected sentences by teachers and  $C_A$  selected ones by the algorithm. Each sentence of every text  $Q$  is defined by its position  $i \in [1, N_Q]$ .  $C[i]$  is the selection of the sentence  $C$  at position  $i$ . Then the similarity for a given algorithm  $A$  can be calculated by the following:

$$sim_A = \frac{1}{Q} \sum_{T=1}^Q \frac{1}{P} \sum_{i=0}^{N_Q} (C_L[i] * C_A[i])$$

$$C[i] = \begin{cases} 1: & \text{if sentence was selected at position } i \\ 0: & \text{else} \end{cases}$$

In our study, we set  $Q = 48$ , and  $P = 3$  as we used 48 texts, 3 selected sentences. The algorithm that is mostly part of the top 1 or top 2 ranking will be selected to be the best performing one.

### 3.3 Human Evaluation

To understand, whether the best performing algorithm can be used practically, we compared the best performing algorithm  $A$  with the known summarizer “LexRank” as  $B$ , which performs best according to the literature [CYG19]. Teachers got the selections of the two algorithms for each text to decide which selection is the most appropriate one for reading comprehension. The positions of both algorithms  $A$  and  $B$  were randomized during the study. Additionally, they had to select single sentences of the algorithmic selections that are practically unusable. Each text was rated by 13 teachers to find the most appropriate one. Fig. 1 visualizes the approach. We used “sumy” [Be19] to summarize all texts, limited to extract 3 sentences only. For Edmundson, we used the texts’ titles as “bonus words” and a list of 566 stop words as additional parameters. For the MMR summarization, we used an existing implementation<sup>4</sup> and adjusted it to get a summary not limited by the number of characters, but for the number of sentences.

<sup>4</sup> [https://github.com/vishnu45/NLP-Extractive-NEWS-summarization-using-MMR/blob/master/mmr\\_summarizer.py](https://github.com/vishnu45/NLP-Extractive-NEWS-summarization-using-MMR/blob/master/mmr_summarizer.py)

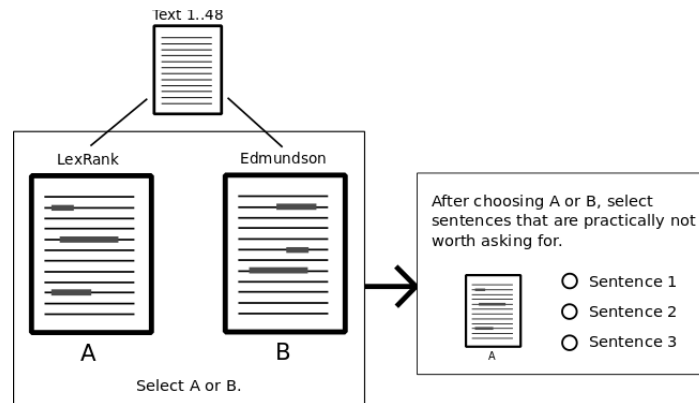


Fig. 1: Study Design to compare selections of two algorithms.

## 4 Results

The pre-study explores the similarity of annotations between teachers and the text summarizers. Each text was annotated by three teachers and we reached an inter-rater-reliability score of 0.8 on average. Tab. 1 shows which algorithm creates the most similar selections compared with teacher ones, averaged over all texts as a ranked list. It shows whether the algorithm is the best performing one (Top 1) or the algorithm is at least one of the two most similar ones compared to teacher selections (Top 2). The summarizer “Edmundson” performs best which answers our 1<sup>st</sup> research question.

Text Summarizer	in Top 1	in Top 2
Edmundson	0.50	0.61
TextRank	0.11	0.39
LexRank	0	0.33
SumBasic	0	0.22
Luhn	0.11	0.22
LSA	0.06	0.17
Longest Sentences	0	0.33
KL	0.17	0.22
MMR	0.06	0.17

Tab. 1: Text summarizers ranked in pre-study.

According to the literature [CYG19], where LexRank performs best, we cannot confirm this result. In our results, LexRank even not occurs in the Top 1. Thus we aimed to compare the two summarizers LexRank and Edmundson in our main study. We let 30 teachers decide which selections they prefer in a practical setting. Each teacher had to choose the

best sentence selection of 16 texts, without telling them that the selections were done by algorithms to avoid prejudice. The same 48 texts of the pre-study were used. We compared the average amount of selections. To find the “winner”, we defined three classes: Edmundson, LexRank, and Ambiguous. The Ambiguous class was chosen if the difference between selections of Edmundson and LexRank was too small. If at least 60% have chosen Edmundson or LexRank, then this was counted to be the preferred version. 37.5% of the selections were ambiguous. By observing the unambiguous selections only, Edmundson was chosen in 73.3% and LexRank was chosen in 26.7%. Thus, Edmundson outperformed LexRank by factor 4.

In addition to the selection of the favored algorithm, all participants were asked to select sentences that they would not use in a practical setting to ask questions about. Teachers selected one of three sentences both for Edmundson and LexRank. This helps to examine the usefulness of using the observed summarizers in real-world scenarios and shows that the summarizer can assist in selecting sentences, but the process still requires an instructor. A complete automatism cannot be achieved at scale yet.

## 5 Limitations

We limited our study to 48 texts, containing different subjects at school. The base for all texts was Wikipedia and all texts were transformed by instructors to be appropriate for students. To achieve more accurate results, the dataset should be extended to use more texts. Also, texts of publishing houses that are broadly used at schools could be used, but these texts cannot be published and studies cannot be done online due to copyright reasons.

For existing datasets, it is important to publish the purpose, why questions were asked. If the purpose is not known, it is difficult to determine whether a selected sentence is good or not. It is not sufficient that selections were just done by instructors or teachers. Without having access to the specific aim of these questions, it is practically not usable. Other didactical concepts could result in other conclusions, too. Further investigations should consider specifying the purpose as close as possible. Furthermore, we have concentrated on the single sentence selection for reading comprehension, because current question generators can only create questions for single sentences. However, if we want to be even more realistic about reading comprehension, we should be able to ask questions that span sentences for more complex questions [B165]. Since this is a much more complex task, we have limited ourselves to single sentences as the first step to this bigger picture.

Text comprehension at school for native speakers is only one concept of using texts. Other methods with different learning goals could be used to extract sentences, that are worth asking for, e.g. in language learning. The resulting optimal algorithm can be different from our findings in other environments. It is important to note that our findings are limited to text comprehension at school. Our result should be compared to other target groups.

## 6 Discussion

Comparing LexRank and Edmundson, Edmundson uses stigmas words, stop words; title and heading (as “bonus words”); and structural indicators [Ed69]. The algorithm filters uninformative words that contain low semantic information. The main criteria for sentence selection are the frequency of words, cue phrases (that contain words like “significant” or “hardly”), heading words (e.g. the title), and the location of sentences within the text. All these features were used to get a score for each sentence. It was found that resulting summaries overlap with 44% of handcrafted ones. In our study, we also used the title as “bonus words” as suggested by Edmundson. In contrast, LexRank [ER04] is an unsupervised method that finds the most important sentences of a given text by using the concept of sentence salience. It uses a graph-based centrality scoring approach to calculate the importance. All sentences’ similarities can be calculated by the eigenvector centrality.

The success of selecting appropriate sentences by Edmundson depends on the input parameters. If all parameters are empty, the algorithm just selects the first and last sentences of a given text. Thus, it is important to use stop and bonus words. For an automatic approach, it is necessary to have access to the header of a text to be used as bonus word(s). Otherwise, these words have to be selected manually, which limits the practical usefulness. Other approaches (e.g. LexRank) are not dependent on the existence of bonus words. All text summarizers are independent of the text lengths and the amount of the selected sentences. Thus, our approach can be used for smaller and larger texts as well. The instructor can set a fixed number of sentences that should be selected. There is no optimal number, the extractive algorithms create a ranked list of potentially question-worthy sentences and the instructor can be supported with that information. As texts are often shorter in practical settings, our approach can be used for sentence selection.

Although we used German texts and German annotators, the approach can be used for all other languages (e.g. English) that are supported by the observed text summarizers. The results of selected sentences by all summarizers will be the same as long as all parameters are identical. Edmundson also will select identical sentences across different languages as long as used bonus words of the header appear in the same sentences as before. Using synonyms in the header or text can change the result.

As described in the results of our study, we asked all participants to choose their favored selection. Afterward, they were supposed to select sentences which they would not ask for in an educational setting to test the usefulness in practice. On average, one out of three sentences was chosen to be impractical. Although the description was clear - only select sentences that were impractical and select none, if all make sense - most participants selected one out of three sentences, even if all the options were valid. Thus the algorithm could be more optimal than the suggested one, as the majority of our participants selected one sentence to be the worst of three, but not an impractical one.



By asking teachers, which concept they used to select sentences that are worth asking for, concerning reading comprehension in secondary education, only 13.5% said they used a concept in our pre-study. All others have chosen sentences by feeling, what they would ask during classes. This brings us to the conclusion, that they may have selected sentences from the texts more subjectively according to their feelings or that the teachers were working from an implicit understanding of what is important. Experts hold considerable implicit knowledge that they are not explicitly aware they possess. The reasons why teachers have mentioned, that they were not using a concept, can be manifold and need to be further investigated. Teachers that used a concept stated, that they selected sentences that were spread over the text, sentences that contained basic information, definitions, to use the Five Ws [SS11], or to concentrate on different taxonomy levels [B165].

Furthermore, it can be discussed, whether there is a bias in the pre-study reasoned at the limitation to exactly 3 sentences since the information content can vary depending on the text. If a text has a very high informational content, it might be difficult for the teachers to be limited to select exactly 3 sentences, because there might be other sentences that have a similar relevance. It is, therefore, possible that there were, for example, 5 relevant sentences in a text, and the teachers also found all 5 important ones but set different priorities in the final decision. If a text had fewer than 3 relevant sentences to which questions could be asked, the teacher would have had to decide on other less relevant sentences by setting the number to 3. Personal priority would also be a factor in this variant. Furthermore, we randomized topics and teachers to get a general view of question-worthy sentences, independent from subjects.

A gap in the training of new teachers could be a reason for missing concepts. If the teachers have not had the opportunity to learn methods and concepts on how to extract important information from texts in a pedagogical context, how to summarize them or how to formulate appropriate questions for text comprehension based on suitable sentences during their training, then these cannot be used by them, as they do not know any.

Another interesting question in this context is whether this gap in qualification between teachers is a general problem or a subject-specific one, since, for example, a teacher of mathematics is comparatively less preoccupied with the text comprehension or extracting the most important information from texts within a pedagogical context than a teacher of language subjects. This does not mean, however, that reading comprehension is not necessary for STEM education, even though this may not be as important as in other subjects. However, even in STEM subjects, longer texts are existing, e.g. on historical or biographical backgrounds. Based on our data, we could only conclude that there was no difference between the teachers' subjects in terms of the use of a concept or method when selecting a sentence. This could be an indicator that neither concepts nor methods were taught to teachers, regardless of the subject.

## 7 Future Work

A personalized way of finding the sentence extractor which is most similar to the teacher selections can help to find sentences that are worth asking for – based on individual preferences. Therefore, a learning period is required, where annotations of selected questions were collected and compared concerning the summarizer extractions. This methodology can be a first step to generate interactive online course material based on the teaching attributes of individual teachers. This approach can help to create a variety of interactive material that can be different across teachers.

By observing the results, we can say that Edmundson performs better than LexRank in the educational context, but for a practical setting, it is not good enough to be used on a large scale. Edmundson uses semantic information based on the topic and the location of sentences [Ed69]. LexRank is a graph-based centrality scoring but focuses on the frequency and deviation of words, not the content itself. As the best performing algorithm of our study uses semantic features, further investigations should focus on semantic approaches as well.

Currently, no algorithm selects all sentences that teachers will choose. But algorithms like Edmundson can support instructors at finding possibly question-worthy sentences. As we were not able to find the most practical algorithm in 37.5% of all cases, a combination of both text summarizers, Edmundson, and LexRank, could be a solution. Further investigations could observe what the best ratio concerning both algorithms will be to get an optimal result. Additionally, the methodology of Edmundson could be used and adjusted to select sentences that are more appropriate for a learning environment.

Due to a limited budget, we decided to compare LexRank, which was the best performing choice according to the literature with Edmundson, that performed best in our case. A comparison of more pairs of summarizers is required to understand whether these both algorithms are the most appropriate ones for educational purposes. However, if teachers are familiar with methods and concepts for selecting sentences that are worth asking questions about, it is still exciting to investigate why they are not used in practice and in which situations they are more suitable for. Teachers may prefer to follow their feelings and expertise, which they have acquired through experience in the field. Since most of the teachers participating in the study stated that they did not use any concept or specific method in selecting the appropriate sentences, it is still interesting to know what criteria they used to select the sentences otherwise. If we know more about the reasons for the selection of the sentences and why it is worth asking questions about them, suitable algorithms could be adapted even more.

During informal talks after the study, we observed that teachers in their training (in Germany) do not learn how and for which purpose to formulate questions, but rather how to formulate tasks with the help of operators. Operators are formulated as verbs (e.g. describe, represent, explain, compare, ...) and indicate what a student should do in concrete

terms. Especially in exam situations teachers should avoid asking questions and instead use operators as keywords because they define specific requirements for students and there can be no misunderstandings as operators are fixed. This can help for further studies to investigate automatic task generation by using operators, instead of questions.

## 8 Conclusion

In this paper, we explored how text summarizers perform with the task of question-worthy sentence selection in an educational environment for text comprehension at school. Therefore, we compared the sentence selections of teachers to the summarizers' ones. The observation of the overlap with the algorithmic choices has shown that the text summarizer "Edmundson" performed best in our setting. We compared this summarizer to LexRank, which was known to be the best performing one in educational settings [CYG19]. The direct comparison of the summarizers' selections has shown that Edmundson was favored by teachers in contrast to LexRank. To examine the usefulness, we observed which selections are not worth to ask questions about. Although the results of applying extractive summarizers to select sentences that are worth asking for in an educational setting are quite promising, one of three selections was marked to be not useful. This shows that a subset of the algorithmic selections can be useful. The preselection can be helpful to support teachers during the selection process and to reduce the time required. From the practical point of view, a fully automatic selection at scale still requires an instructor.

**Acknowledgments:** This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII116 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors.

## Bibliography

- [Be19] Belica, M.: Automatic text summarizer: sumy 0.8.1, 19 05 2019. [Online]. Available: <https://pypi.org/project/sumy/>. [Accessed 12 12 2018].
- [Bl65] Bloom, B.: Bloom's Taxonomy of educational objectives, in Vol. 1: Cognitive domain, New York, 1965.
- [Ch18] Chen, G. et al.: LearningQ: A Large-scale Dataset for Educational Question Generation, in AAAI, 2018.
- [CYG19] Chen, G.; Yang, J.; Gasevic, D.: A Comparative Study on Question-Worthy Sentence Selection Strategies for Educational Question Generation, in AIED, Chicago, 2019, pp. 59-70.
- [DC17] Du, X.; Cardie, C.: Identifying Where to Focus in Reading Comprehension for Neural Question Generation, in Proceedings of the Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, Association for Computational Linguistics, 2017, p. 2067–2073.

- 
- [De18] Devlin, J. et al.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018.
- [De90] Deerwester, S. C. et al.: Indexing by latent semantic analysis, in *JASIS* 41, 6, 1990, p. 391–407.
- [DSC17] Du, X.; Shao, J.; Cardie, C.: Learning to Ask: Neural Question Generation for Reading Comprehension, in *ACL*, 2017.
- [Ed69] Edmundson, H. P.: New methods in automatic extracting, in *Journal of the Association for Computing Machinery*, Vol. 16, No. 2., 1969, pp. 264-284.
- [ER04] Erkan, G.; Radev, D. R.: LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization, in *Journal of Artificial Intelligence Research* 22, AI Access Foundation, 2004, pp. 457-479.
- [GC98] Goldstein, J.; Carbonell, J.: Summarization: (1) Using MMR for Diversity-based ReRanking and (2) Evaluating Summaries, in *Tipster Text Program Phase III: Proceedings of a workshop*, Baltimore, Maryland, USA, Association for Computational Linguistics, 1998, pp. 181-195.
- [HV09] Haghighi, A.; Vanderwende, L.: Exploring Content Models for Multi-Document Summarization, in *Annual Conference of the North American Chapter of the ACL*, Colorado, Association for Computational Linguistics, 2009, pp. 362-370.
- [HS16] Heilman, M.; Smith, N. A.: Question Generation via Overgenerating Transformations and Ranking, Pittsburgh, Language Technologies Institute, 2009.
- [Jo17] Joshi, M. et al.: TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension, in *Computation and Language*, CoRR, 2017.
- [Li04] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, in *Text Summarization Branches Out*, Association for Computational Linguistics, 2004, p. 74–81.
- [Lu58] Luhn, H. P.: The Automatic Creation of Literature Abstracts, in *IRE National Convention*, New York, IBM Journal of Research and Development, 1958, pp. 159-165.
- [MT04] Mihalcea, R.; Tarau, P.: TextRank: Bringing Order to Texts, in *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Spain, Association for Computational Linguistics, 2004, p. 404–411.
- [NV05] Nenkova, A.; Vanderwende, L.: The Impact of Frequency on Summarization, 2005.
- [Pa02] Papineni, K. et al.: BLEU: A method for automatic evaluation of machine translation, in *Association for Computational Linguistics*, Philadelphia, Pennsylvania, ACM, 2002, pp. 311-318.
- [Ra16] Rajpurkar, P. et al.: SQuAD: 100,000+ Questions for Machine Comprehension of Text, in *Conference on Empirical Methods in Natural Language Processing*, 2016.
- [SS11] Salgado, S.; Strömbäck, J.: Interpretive journalism: A review of concepts, operationalizations and key findings, *Sage Journals*, 2011.

## Is the context-based Word2Vec representation useful to determine Question Words for Generators?

Sylvio Rüdian<sup>1</sup> and Niels Pinkwart<sup>2</sup>

**Abstract:** Question and answer generation approaches focus on the quality and correctness of generated questions for online courses but miss to use a good question word, which is a deficiency reported by many previous studies. In this experimental setup, we explored whether the word2vec representation, which is semantic-based, can be used to predict question words. We compare two pipelines of the prediction process and observed that splitting the problem into several subproblems performs similar to feeding a neural network with all the data. Although our approach is promising to take the context-based representation into account we can see that the success rate is still low but better than guessing.

**Keywords:** Question generation; question words, online courses;

### 1 Introduction

Question generation for creating interactive learning material in online courses has been investigated a lot within the last decade. The general idea is to take a text and an algorithm has the task to generate appropriate questions that can be used in online courses. While the process of generating questions by patterns or templates works well, most of these approaches work on the syntactical basis only, but they miss the semantics.

By observing texts and single sentences where teachers can ask questions about, we always take the context into account. If a text is about a location, a question using the question-word (WH-word): “Where” can often be used. When asking about persons, “Who” can be used, etc. Data representations like word2vec use semantics and relations of words to each other which creates many clusters of connected words. Entities that are connected have a low distance in the graph. Our idea is to combine this cluster with already known WH-words. If we know that a keyword is part of a special cluster of the word2vec structure (e.g. a place) and we already know typical question words (like “Where”) or templates (like “Where is [X] located in?”) then we can use this to generate a question without the necessity to determine manually that we are talking about a place.

---

<sup>1</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin, Weizenbaum Institute for the Networked Society, ruediasy@informatik.hu-berlin.de, <https://orcid.org/0000-0003-3943-4802>

<sup>2</sup> Humboldt Universität zu Berlin, Institut für Informatik, Unter den Linden 6, 10099 Berlin  
pinkwart@hu-berlin.de

## 2 Related Work

Heilman et al. [HS09] introduced an approach of question generation via overgenerating, transformations, and ranking. They used sentences as inputs and generated multiple questions following a rule-based approach. Therefore, they manually improved rules and expressions applied manual conditions and defined 12 features to rank generated questions with a supervised approach. This approach is based on a large, manually crafted set of rules that can be used to generate questions.

Zhao et al. [Zh11] used queries to generate questions for “community-based question answering”. The idea of the authors was that users have a question in mind when searching for specific information. The generation of a question based on a few keywords could help. Therefore, they investigated an approach of automatically generated templates that can be used to create questions. Templates contain placeholders that can be replaced by entities. But templates like “What is [x2] of [x1]” are really specific and cannot generalize well. Finally, a lot of training data is required to create good templates with the right question words. Rodrigues et al. [RCN16] introduced a framework to generate questions based on different levels of linguistic information. They used triples as training data, with each triple consisting of a question, an optional answer, and a snippet that could answer the question. After learning the structure, they used the syntactic tree representation of questions and snippets and created a mapping between subtrees of questions and corresponding snippets. The final model consists of 23 semantic patterns that can be used to generate questions based on textual snippets, but the focus on selecting a good WH-word is missing.

Kantor et al. [KKS14] introduced a pipeline to use grammar patterns of sentences to generate questions and additionally used the “semantic association of a verb” to derive the right question type to find the optimal WH-word for a question. They created a manual corrected question type database and used only verbs to determine which of the WH-word is the correct one. G. Chen et al. [CYG19] used four different datasets (TriviaQA [Jo17], MCTest, RACE, and LearningQ) to examine which existing text summarizers select the optimal sentences which were asked for. But the evaluation of generated questions does not focus on the selection of a good and appropriate question word so that a lot of questions starting with “what”. Rüdian et al. [RP19] introduced a pipeline for teachers to generate questions and corresponding answers for online courses while they can improve the dataset by giving manual corrections in natural language. The results are promising but the focus on choosing an optimal question word is not supported and outlined for further investigations.

In this paper, we explore the problem of wrong question words, which is a deficiency identified by Heilman [HS09]. Kantor et al. [KKS14] limited the selection to use verbs only for finding the best WH-word. We use a combination of verbs, nouns of the question and the answer to decide which is automatically chosen as the best. Therefore, we use the word2vec representation of our focused keywords and examine whether this dataset can be used to predict WH-words.

### 3 Methodology

Our training data consists of triples like in [RCN16], which consist of a sentence, a question, and the corresponding answer. We used the Stanford Question Answering Dataset (SQuAD 1.1 [Ra16]). It consists of structured texts with corresponding questions and answers. We assume that we can use this data to learn which WH-word is appropriate for which question. We extracted all Q&A-pairs and combined them with the related sentences of the texts. In contrast to [RCN16], all learned questions have an answer, because they are existing in the dataset.

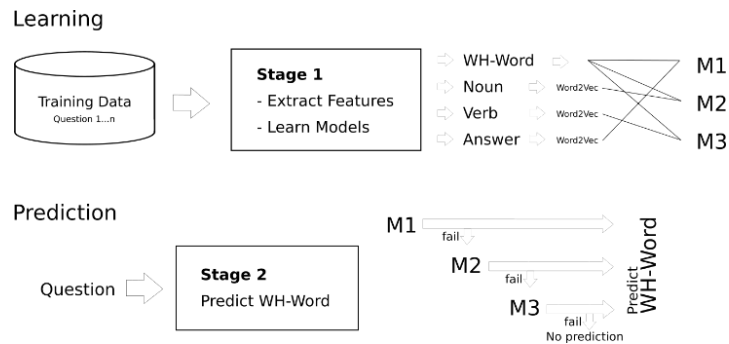


Fig. 1: Process of learning and predicting WH-words.

The word2vec representation of words [MLS13] maps words into a 300-dimensional vector space in which related words have a smaller distance in hyperspace than others. We assume that this representation helps to classify corresponding WH-words since we can often use the same WH-word for similar words. We prepared three different models to be able to compare different results and to have the possibility of fallbacks in case that an error occurs due to missing words where no word2vec representation exists. First, we used the noun of the answer (M1) and created a supervised neural network. Second, we used the noun of the question (M2) that has a minimal amount of words between itself and the WH-word. Third, we used the verb of the question (M3) that has the smallest distance to the WH-word like in the 2<sup>nd</sup> model. In all three models, we used the extracted word and chose the word2vec representation as features plus the known WH-words as labels for training. The approach is visualized in Fig. 1. According to the three models, the approach allows us to predict the optimal WH-word. The prediction can be applied like a pipeline, using the noun of the answer first.

We created two different ways of predicting WH-words because we wanted to test whether predicting single WH-words and compiling the results would perform better than predicting one out of seven. In the first approach, we learn the word2vec representations and put them into a neural network, those labels consist of one of seven WH-words. In our second approach we learn individually for each of the seven WH-words whether it is the best. To predict WH-words, we calculate the accuracy for each WH-words separately. The WH-word with the highest accuracy will be chosen for prediction. To get the accuracy, we use

10-fold-Cross-Validation (10f-CV) as a classical method in machine learning. Therefore, we split our dataset into 10 pieces of the same length. Then we train our model with data of 9 pieces and test it with the remaining one. Thus, we can test predictions on previously unseen data, where we already know the labels and compare them with our predictions. Finally, we can compare both approaches.

## 4 Results

A major problem is that trained patterns include “What” as WH-word mostly, as this question word occurs more frequently than others. One question deficiency of the evaluation by Heilman et al. is the existence of wrong WH-words [HS09]. We can confirm the dominance of the question word “what” (Tab. 1). In our training dataset, 58% of all questions use it. Often “what” is not the best WH-word. Questions starting with “What date...” or “What lives...” could be replaced by using WH-words like “When” or “Where”.

In the following section, we show the results of predicting WH-words, using neural networks. The optimal sequence in which nouns, verbs, or answers are used depends on the availability of their word2vec representations and the final accuracy of each approach. We used the publicly available word2vec representation<sup>3</sup> of Mikolov et al [MLS13]. Training data consists of 10.917 questions with corresponding answers. 92.6% of question nouns are covered by the Word2Vec representation. 58.7% of the verbs can use the word2vec representation and 19.8% of the answer nouns can be found in word2vec. Tab. 1 shows all absolute values where we could convert given words with word2vec representation, separated by WH-words. By combining all three approaches (optimal sequence: M1, M2, M3) to predict the most suitable WH-word as proposed in Fig. 1, we can cover 97.8% of our training data using the word2vec representation. When we use the combination, we cover more training data as opposed to using each of the three approaches alone.

Word	What	Who	When	How	Which	Where	Why
Noun in question (M3)	5908	1081	490	1264	962	327	81
Verb in question (M2)	3689	505	405	908	517	307	73
Noun in answer (M1)	1244	264	113	231	209	83	16

Tab. 1: Word2Vec Coverage of WH-words in the dataset of 10846 samples

We focused on two different approaches to predict WH-words. First, for each of the three words we used a neural network that has to predict the best WH-word (what, who, when, how, which, where, and why). Our second approach asks for each WH-word whether it is the best and we finally use the one with the highest probability among the 7 different neural networks. We wanted to compare these two approaches whether predicting each WH-word first and combining it leads to better accuracy than predicting one WH-word out of seven directly. For each of our approaches, we created a model and optimized the

<sup>3</sup> Available at <https://code.google.com/archive/p/word2vec/>



hyperparameters by using a grid search. In each case, we only use samples that have a word2vec representation.

If we combine all three models in our pipeline, the overall accuracy in 10f-CV reaches 44.9% after hyperparameter optimization, covering 97.8% of our dataset. For our second approach, where we used each WH-word separately, we identified all 5 hyperparameters for individual predictions (M1) to (M3). We calculated the 10f-CV using our previously described pipeline to predict whether a single WH-word is the correct one. The results show that the accuracy of prediction for individual WH-words (66-81%) is much better than our first approach. To find out if it is better to ask for individual WH-words, instead of getting one WH-word of seven, we combined all models with optimized hyperparameters. The prediction of whether a single WH-word is the right one will result in a probability score. The prediction with the highest probability is selected as the output. Interestingly, the result of the mentioned second approach is in the same range as the previous one, it reached 43.2% after hyperparameter optimization. The covering rate is also the same. Thus we can see that it is not necessary to observe single WH-words and combine them to predict the best fitting WH-word. A comprehensive neural network is sufficient.

## 5 Discussion

One general problem of WH-word optimization is imbalanced data. Tab. 1 shows the distribution of available samples, where we can see, that nearly 58% of all questions use the WH-word “What”. If we labeled every prediction with “What”, our accuracy would reach 58%, just because of the distribution. During each training process, we balanced data and removed additional existing samples if it was necessary. Finally, WH-words “When”, “Where” and “Why” were underrepresented, resulting in a general drop in accuracy of nearby 10% each. Our results of the two approaches for prediction show that the accuracy is fundamentally just a little bit better than guessing. Further investigations should focus on obtaining more training data to improve results. If this does not help, the classical approach with using hierarchical databases or WordNet [Fe12] with domain knowledge should be applied in real scenario applications. Besides, it can also be the case that the selection of WH-words does not depend on the context only. Instead, it can depend on other aspects, e.g. the task, that should be taken into account in further investigations.

## 6 Conclusion

In this paper, we explored an approach to overcome the deficiency of wrong WH-words in question generating systems. To do that, we examined whether the word2vec representation can be applied to determine WH-words. Results have shown that this representation can generally be used, but the result of ~45% accuracy is not as good as we would like to be. To propose WH-words it works, though. Finally, it is recommended not to use this data structure only to predict question types. A combination with other features is required,

that have not been used yet, e.g. the type of the task where a question needs to be created. This can help to increase accuracy in further investigations.

Besides we showed that using a single neural network, that contains all existing data, performs similar to using several neural networks, where each has been optimized individually to obtain a summarized result. Training and optimizing hyperparameters of multiple neural networks is not required if the task can be done with a single neural network instead.

**Acknowledgments:** This work was supported by the German Federal Ministry of Education and Research (BMBF), grant number 16DII116 (Weizenbaum-Institute). The responsibility for the content of this publication remains with the authors.

## Bibliography

- [CYG19] Chen, G.; Yang, J; Gasevic, D.: A Comparative Study on Question-Worthy Sentence Selection Strategies for Educational Question Generation, in AIED, Chicago, 2019, pp. 59-70.
- [Fe12] Fellbaum, C.: WordNet, in The Encyclopedia of Applied Linguistics, C. Chapelle (Ed.), 2012.
- [HS09] Heilman, M.; Smith, N. A.: Question Generation via Overgenerating Transformations and Ranking, Pittsburgh, Language Technologies Institute, 2009.
- [Jo17] Joshi, M. et al.: TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension, in Computation and Language, CoRR, 2017.
- [KKS14] Kantor, A.; Kleindienst, J.; Schmid, M.: Automatic question generation from natural text. US Patent US9904675B2, 27 10 2014.
- [MLS13] Mikolov, T.; Le, Q. V.; Sutskever, I.: Exploiting Similarities among Languages for Machine Translation, Mountain View, Google Inc., 2013.
- [Ra16] Rajpurkar, P. et al.: SQuAD: 100,000+ Questions for Machine Comprehension of Text, in Conference on Empirical Methods in Natural Language Processing, 2016.
- [RCN16] Rodrigues, H. P.; Coheur, L.; Nyberg, E.: QGASP: a Framework for Question Generation Based on Different Levels of Linguistic Information, in Proceedings of the 9th International Natural Language Generation conference, Edinburgh, UK, Association for Computational Linguistic, 2016, pp. 242-243.
- [RP19] Rüdian, S.; Pinkwart, N.: Towards an Automatic Q&A Generation for Online Courses - A Pipeline Based Approach, in Artificial Intelligence in Education (AIED 2019), Chicago, Springer, 2019, pp. 237-241.
- [Zh11] Zhao, S. et al.: Automatically Generating Questions from Queries for Community-based Question Answering, in Proceedings of the 5th International Joint Conference on Natural Language Processing, Chiang Mai, Thailand, AFNLP, 2011, p. 929–937.

## Was bedeutet „Digitaler Unterricht“?

### Ein Kriterienkatalog zur Befundung digitaler Medien

Timon Bauer<sup>1</sup>, Sven Hofmann<sup>2</sup>, Stefanie Kersten<sup>3</sup>, Marcel Lehmann<sup>4</sup> und Silvia  
Schöneburg-Lehnert<sup>5</sup>

**Abstract:** Im Rahmen des Projekts UndiMeS (Unterrichten mit digitalen Medien in Sachsen) werden „hoch-digitalisierte Unterrichtsszenarien“ für die Fächer Mathematik und Informatik im Bereich der Sekundarstufe I entwickelt, empirisch erprobt und den Lehrkräften an den Schulen zur Nachnutzung zur Verfügung gestellt.

Um Aussagen treffen zu können, unter welchen Bedingungen Lehr-Lern-Szenarien tatsächlich als „digitalisiert“ zu bezeichnen sind, wurde ein Kriterienkatalog entwickelt, der bereits praktizierte Unterrichtsszenarien mit dem Fokus auf den Digitalisierungsgrad befunden lässt und als Richtlinie für die Entwicklung neuer Lehr-Lern-Szenarien dienen kann. Dieser Katalog wird stetig weiter ausdifferenziert und fokussiert das digitale Medium in der Mikroebene. Die resultierenden Kriterien werden nun mit deren Arbeitsstand präsentiert und zur Diskussion gestellt.

**Keywords:** Digitalisierung, Digitaler Unterricht, Webbasiertes Lernen, Digital unterstütztes Lehren und Lernen

## 1 Einleitung

Digitalisierung, webbasiertes Lernen und digitale Bildung zählen noch immer zu den Schlüsselwörtern unserer Zeit. Mehr als die Hälfte aller Lehrkräfte würden die digitalen Medien gern häufiger einsetzen, scheitern aber an verschiedenen Hürden [bi19, S. 6]. So liegt beispielsweise die Relation zwischen den Lernenden und digitalen Devices mit 9,7:1 in Deutschland unter dem Durchschnitt, welchen die ICILS-Studie 2018 veröffentlichte [EBG19, S. 14]. Dies lässt vermuten, dass bei einer durchschnittlichen Verfügbarkeit eines Geräts für ca. zehn Lernende eine flexible Durchführung digitaler Lehr-Lern-Szenarien nicht möglich ist.

---

<sup>1</sup> Universität Leipzig, Professur für Didaktik der Mathematik, Augustusplatz 10, 04109 Leipzig,  
timon.bauer@math.uni-leipzig.de

<sup>2</sup> Universität Leipzig, Professur für Didaktik der Informatik, Augustusplatz 10, 04109 Leipzig,  
sven.hofmann@informatik.uni-leipzig.de

<sup>3</sup> Universität Leipzig, Professur für Didaktik der Mathematik, Augustusplatz 10, 04109 Leipzig,  
stefanie.kersten@math.uni-leipzig.de

<sup>4</sup> Universität Leipzig, Professur für Didaktik der Informatik, Augustusplatz 10, 04109 Leipzig,  
marcel.lehmann@uni-leipzig.de

<sup>5</sup> Universität Leipzig, Professur für Didaktik der Mathematik, Augustusplatz 10, 04109 Leipzig,  
silvia.schoeneburg@math.uni-leipzig.de

Aber was heißt “digitaler Unterricht”? Wann gilt ein Lehr-Lern-Szenarium tatsächlich als “digitalisiert”? Und wie sollten entsprechende Medien entwickelt sein?

Zur Beantwortung dieser Fragen wird ein Messinstrument zur Kategorisierung digitalen Lehrens und Lernens – ein Kriterienkatalog – benötigt, der messbare, objektive Bewertungskriterien bietet und eine Leitlinie für Lehrkräfte darstellt, um die Anforderungen an digital gestützte Unterrichtssettings nachvollziehen und auf den eigenen Unterricht anwenden zu können. Insbesondere soll der Katalog im Rahmen einer wissenschaftlichen Begleitforschung ermöglichen, die Qualität von Medien auf den Unterricht zu überprüfen und so zu untersuchen, ob ein Einfluss auf den Lernerfolg vorhanden ist. Aufgrund dessen wird der Katalog anwendungswissenschaftlich angelegt [DB16, S. 186].

## **2 Digitale Medien als Fundament digitaler Lehr-Lern-Szenarien**

Als Medium wird im Kontext von Lehr- und Lernprozessen ein Mittler bezeichnet, der in kommunikativen Zusammenhängen potenzielle Zeichen mit technischer Unterstützung speichert, wiedergibt, anordnet oder verarbeitet; und in bildhafter und/oder symbolischer Form präsentiert [HG06, S. 10].

Digitale Medien sind computerbasierte Angebote von Informations- und Kommunikationstechnologien mit den entscheidenden Funktionen der Verarbeitung, Speicherung, Verteilung, Darstellung, Berechnung und Aufzeichnung von Daten [HG06, S. 11]. Im Vergleich zu den traditionellen werden die digitalen Medien dynamischer bezüglich der Interaktivität und Handlungsfähigkeit der Nutzenden.

Die Wahl des digitalen Mediums wird vom Unterrichtsprozess determiniert. Die Lehrkraft muss dafür verschiedene didaktische Entscheidungen treffen, die sich nicht in einem einfachen eindimensionalen Ursache-Wirkungs-Prinzip wiederfinden. Einflussfaktoren wirken sowohl seitens der Lernenden (z.B. Vorwissen), der Lehrperson (z.B. fachdidaktische Expertise) und des Unterrichtsprozesses (z.B. Sozialform), als auch seitens des Mediums (z.B. Interaktivität) [He14, S. 10]. Es ist daher falsch, anzunehmen, dass digitale Medien automatisch Lernerfolge bei den Lernenden auslösen. Jede Anwendung im Unterricht ist ein hochkomplexer Prozess, der seitens der Akteure und des Prozesses selbst unterschiedliche Rahmenbedingungen enthält. Alle Wechselwirkungen der vier Dimensionen - Lehrkraft, Lernende, digitales Medium und Unterrichtsprozess - sind entscheidend für die Wirkung im Lehr-Lern-Prozess. Daraus lassen sich drei Folgerungen ableiten, die den weiteren Diskurs beeinflussen [HL20, S. 5562]:

1. Der Einsatz digitaler Medien ist essenziell für die Gestaltung digitalen Unterrichts.

2. Der Einsatz kann zwar notwendig, allerdings nicht hinreichend für guten digitalen Unterricht sein. Dies erfordert stets eine didaktische, fachliche und technologische Komponente.
3. Durch den komplexen Wirkungskreislauf lassen empirische Zugänge zur Beurteilung von digitalen Medien, und damit von digitalem Unterricht, keine pauschalen Schlüsse zu.

Zusammenfassend muss eine Vielfalt an Kriterien entstehen, um hoch-digitalisierte Unterrichtsszenarien mit deren digitalen Medien zu bewerten und anschließend zu nutzen. Dafür ist eine Orientierung an den theoretischen Aspekten von digitalen Medien und deren Einflussfaktoren nötig.

### **3 Entwicklung des Kriterienkataloges**

Ein im Rahmen einer Staatsexamensarbeit als Vorarbeit bereits entwickelter Basiskatalog für digitalen Unterricht [HL20] und eine systematische Literaturrecherche (u.a. [Wa18, Ur14, Ur10, Dr19]) stellen die Grundlage der Kriterienentwicklung dar. Neben einigen wenigen Quellen, welche den Einsatz digitaler Medien in der Sekundarstufe I an deutschen Oberschulen und Gymnasien fokussieren, wurde vor allem auf entsprechende Literatur aus dem grundschuldidaktischen Bereich zurückgegriffen, die Kriterien anschließend aus häufig auftretenden Potenzialen, bzgl. des digitalen Unterrichts, analysiert und für den Katalog deduktiv abgeleitet und entsprechend für Oberschule und Gymnasium angepasst implementiert.

Durch dieses Vorgehen wurde eine Reihe von fachlichen, z.B. Aufgabengestaltung [Kr12, S. 45], technologischen, z.B. Darstellungswechsel [Wa18, 43ff], und didaktischen Kriterien entwickelt, wie bspw. das Schrankenkriterium [Sc06, 205ff]. Letzteres hat im Katalog einen großen Stellenwert, da es explizit den Mehrwert von digitalen Medien gegenüber den analogen Gegenständen messen soll.

Somit werden möglichst vielfältig und umfassend die einzelnen Facetten der Fragestellung, wie digitale Medien und digitaler Unterricht zu optimalen Lernerfolgen bei den Schülerinnen und Schülern führen, abgedeckt. Ebenso werden Kriterien zum allgemeinen Setup des Unterrichts untersucht und in den Katalog integriert. Hierzu gehören u.a. Aspekte des zugrundeliegenden Lernmodells und Eigenschaften von selbstreguliertem Lernen. So wird empfohlen, eine konstruktivistische Haltung einzunehmen und den Fokus vor allem auf das selbstständige Lernen zu legen [Ro19, S. 240]. Dadurch können beispielsweise Potenziale zur Veränderung von Klassenraumdynamiken und zur Weiterentwicklung des metakognitiven Verständnisses verwirklicht werden [BS19, S. 122].

Zusätzlich zu den deduktiv abgeleiteten Kriterien existieren auch diese, welche aus pragmatischen, schulpraxisrelevanten Gründen eingefügt werden. Dies betrifft u.a. das Kriterium „Datenschutz“. Aufgrund dieses Vorgehens ist eine empirische Überprüfung

des bisher theoriegeleiteten Katalogs unabdingbar. Dafür soll in einem der nächsten Schritte vor allem in Zusammenarbeit mit Lehrkräften erfolgen.

#### 4 Kriterienkatalog für digitale Medien

Auf der Grundlage dieser Vorgehensweise nimmt der Katalog sowohl einen deskriptiven als auch explanativen Charakter an. Er beinhaltet dementsprechend latente Kriterien, um Vorhersagen über den Lernerfolg treffen zu können [DB16, S. 192]. Im Folgenden wird ein Ausschnitt des Katalogs präsentiert mit jeweils zwei Repräsentanten jeder Dimension.

Kriterium	Beschreibung
<b>Technologische Dimension</b>	
Darstellungswechsel	Das digitale Medium stellt den fachlichen Inhalt auf allen drei Darstellungsebenen des Brunerschen EIS-Prinzips dar und ermöglicht eine Verknüpfung der Ebenen miteinander.
Mehrperspektivität	Durch die Anwendung der Designprinzipien Kontiguität, Multimedia, Kohärenz und Redundanz auf das digitale Medium wird ein multimodales (mehrkanaliges) und multicodales Lernen ermöglicht.
<b>Fachliche Dimension</b>	
Kompetenzkriterium	Die Handlung mit dem Medium und dessen Inhalte fördern fach- und methodenspezifische Kompetenzen.
Fachliche Zielsetzung	Das Medium verfolgt ein klar festgelegtes thematisches/fachliches Ziel.
<b>Didaktische Dimension</b>	
Schrankenkriterium	Die effektive und effiziente Nutzung des digitalen Mediums ermöglicht das Überwinden der Zeit - und Raumschranke des E-Learning, expandiert die Lernchancen und sichert einen Mehrwert gegenüber analogen Alternativen.
Förderliche Lernumgebung	Die Verwendung des digitalen Mediums ermöglicht eine für digitalen Unterricht angemessene Lernumgebung, in der ein selbstständiger und interaktiver Lernprozess eröffnet wird.

Tab. 1 Tabellarischer Kriterienkatalog für digitale Medien (Auszug)

Für die einzelnen Kriterien erfolgt als nächstes ein erster Vorschlag einer schrittweisen Graduierung, wie in nachfolgender Tabelle 2 für das Kriterium „Darstellungswechsel“ zu sehen ist. Basierend auf der Vorarbeit [HL20, S. 5564] besteht diese aus vier aufeinander aufbauenden Stufen, wobei Stufe 3 die Zielstufe beschreibt. Das vierstufige System ist vorteilhaft, um die Tendenz zur Mitte zu vermeiden und eine präzisere Bewertung zu erreichen [BD15, S. 184]. Die Graduierung entspricht somit zunächst einer Ordinalskala. Sie soll als Werkzeug zur besseren Evaluation der digitalen Medien dienen und eine

gewisse Trennschärfe in den Umsetzungsniveaus der Kriterien ermöglichen. In diesem Zusammenhang ist es auch denkbar, die bestehende Graduierung um weitere Stufen zu erweitern, um die Trennschärfe eindeutiger festlegen zu können.

<b>Grad</b>	<b>Beschreibung</b>
Stufe 0	Das digitale Medium enthält keine unterschiedlichen Darstellungsebenen. Ein Wechsel ist daher nicht möglich.
Stufe 1	Der Wechsel der Darstellungsebenen ist in geringem Maße zwischen höchstens zwei Ebenen möglich.
Stufe 2	Grundsätzlich wurden alle Darstellungsebenen auf das digitale Medium übertragen. Der Wechsel zwischen zwei Ebenen ist möglich.
Stufe 3	Das EIS-Prinzip wurde vollständig in das CEIS-Prinzip überführt. Das digitale Medium lässt einen Wechsel der drei Darstellungsebenen im Lernprozess zu.

Tab. 2 Graduierungsstufen für das Kriterium „Darstellungswechsel“

## 5 Ausblick

Auf der Grundlage dieses entwickelten und im Folgenden auch noch weiter auszudifferenzierenden Kriterienkatalogs sollen künftig bereits bestehende digitale Szenarien untersucht und beurteilt werden. Ziel ist es, danach selbst solche digitalen Szenarien zu erstellen und diese anschließend mit Lernenden und Lehrkräften zu erproben und zu evaluieren. Die Evaluation ist als Mittel der empirischen Verifikation im Rahmen mehrerer Laborstudien und später auch Feldstudien angedacht, in denen der Kriterienkatalog einem Usability-Test unterzogen wird. Da die Teilnehmenden aufgrund der fest zusammengesetzten Schulklassen nicht randomisiert werden können, bietet sich ein quasi-experimentelles Untersuchungs-Set-Up an [DB16, S.207]. Es ist zu berücksichtigen, dass der Kriterienkatalog nach dieser ersten Projektphase keinen Anspruch auf Vollständigkeit erhebt. Vielmehr ist die Evaluation der Medien, der Lehr-Lernumgebungen und des Katalogs im Sinne einer kontinuierlichen Verbesserung zu verstehen [Eh17, S. 22]. Zwecks dessen soll der Kriterienkatalog nach weiterer Ausreifung an Lehrkräfte weitergegeben werden, wodurch aufgrund des praxisnahen Einsatzes eine stetig wachsende Validität erwartet wird.

## Literaturverzeichnis

- [BD15] Bortz, J.; Döring, N.: Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler. Springer, Berlin, 2015.
- [bi19] bitkom: Smart School – Auf dem Weg zur digitalen Schule. Smart School –Auf dem Weg zur digitalen Schule, 27.03.2020.

- [BS19] Ball, S.; Stacey, K.: Technology-supported classrooms: New opportunities for communication and development of mathematical understanding. In (Büchter, A.; Glade, M.; Herold-Blasius, R. Hrsg.): *Vielfältige Zugänge zum Mathematikunterricht. Konzepte und Beispiele aus Forschung und Praxis*, 2019; S. 121–131.
- [DB16] Döring, N.; Bortz, J.: *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer, Berlin, Heidelberg, 2016.
- [Dr19] Drijvers, P.: Head in the clouds, feet in the ground – A realistic view on using digital tools in mathematics education. In (Büchter, A.; Glade, M.; Herold-Blasius, R. Hrsg.): *Vielfältige Zugänge zum Mathematikunterricht. Konzepte und Beispiele aus Forschung und Praxis*, 2019; S. 163–177.
- [EBG19] Eickelmann, B.; Bos, W.; Gerick, J.: ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking, 2019.
- [Eh17] Ehlers, U.-D.: *Evaluation von E-Learning. Checklisten, Kriterienkataloge oder Evaluationskonzepte? Zum Stand der Bewertungsverfahren für E-Learning-Arrangements*.
- [He14] Herzig, B.: *Wie wirksam sind digitale Medien im Unterricht?*, 2014.
- [HG06] Herzig, B.; Grafe, S.: *DIGITALE MEDIEN IN DER SCHULE. STANDORTBESTIMMUNG UND HANDLUNGSEMPFEHLUNGEN FÜR DIE ZUKUNFT*, 2006.
- [HL20] Hofmann, S.; Lehmann, M.: WHAT DO DIGITAL LESSONS MEAN? In (IATED Academy Hrsg.): *INTED 2020. Conference proceedings*. IATED Academy, Valencia, 2020; S. 5558–5565.
- [Kr12] Krauthausen, G.: *Digitale Medien im Mathematikunterricht der Grundschule*. Spektrum Akademischer Verlag, Berlin, 2012.
- [Ro19] Roth, J.: Digitale Werkzeuge im Mathematikunterricht – Konzepte, empirische Ergebnisse und Desiderate. In (Büchter, A.; Glade, M.; Herold-Blasius, R. Hrsg.): *Vielfältige Zugänge zum Mathematikunterricht. Konzepte und Beispiele aus Forschung und Praxis*, 2019; S. 233–249.
- [Sc06] Schulmeister, R.: *eLearning: Einsichten und Aussichten*. Oldenbourg, München, 2006.
- [Ur10] Urff, C.: Potentiale und Perspektiven digitaler Lernmedien für die Förderung grundlegender mathematischer Kompetenzen. In *Zeitschrift für Heilpädagogik*, 2010, 61; S. 141–150.
- [Ur14] Urff, C.: *Digitale Lernmedien zur Förderung grundlegender mathematischer Kompetenzen. Theoretische Analysen, empirische Fallstudien und praktische Umsetzung anhand der Entwicklung virtueller Arbeitsmittel*. Zugl.: Ludwigsburg, Pädagogische Hochschule, Diss., 2013. Mensch und Buch Verl., Berlin, 2014.
- [Wa18] Walter, D.: *Nutzungsweisen bei der Verwendung von Tablet-Apps*. Dissertation, 2018.



## Mobile First: A Trend in Virtual Learning Environments

Shirin Riazzy<sup>1</sup>, Katharina Simbeck<sup>2</sup>, Marco Träger<sup>3</sup>, Robert Wöstenfeld<sup>4</sup>

**Abstract:** Although mobile learning has long been predicted to become a vital part of the educational reality, schools often seem reluctant to implement mobile teaching solutions. In order to assess the current preferences of learning modalities for school students (ages 9-15) and teachers, an e-learning environments traffic data was analyzed. We have detected two trends: The first is a total rise of mobile usage, especially in comparison to the usage of desktop PCs in the past four years. Second, we have detected that especially students aged 12-15 mostly prefer mobile devices. Hence platform design should be adapted for better use with mobile devices to meet the learners' needs.

**Keywords:** Mobile devices, MOOC, Google Analytics, Attrition.

### 1 Introduction

With over 50% of all online page views being performed on mobile devices [St19], a primary focus of website design has shifted to from desktop PCs to mobile devices [PZ15]. The *mobile first* design is a philosophy in user interface design for websites, which states that the design should be created first (and optimized) for mobile devices [Mu15]. This philosophy has since been extended to responsive web design, where the platform, screen size and, recently, contexts adapt to the users' behavior [Mu15, PZ15]. Other solutions to the rising mobile trend, such as adaptive web designs or separate sites for mobile devices have also been implemented [CP17].

This trend is also developing in e-learning contexts, with a rise in mobile learning applications and websites [Wa15]. In current research, the technology acceptance model is often used to evaluate views of students towards so-called m-learning, meaning learning through a portable device, often smartphones or tablets. However, many researchers define m-learning in a broader sense, incorporating the mobility of technologies, the mobility of learners, of educators and of learning itself [AES16].

In the examination at hand, user data from a German virtual learning platform (Mathe im Advent, short: MiA) was evaluated. MiA offers an annual prize competition in the form of an advent calendar for students who correctly solve 24 mathematical tasks on 24

---

<sup>1</sup> University of Applied Science, Treskowallee 8, 10318 Berlin, Germany, riazzy@htw-berlin.de

<sup>2</sup> University of Applied Science, Treskowallee 8, 10318 Berlin, Germany, simbeck@htw-berlin.de

<sup>3</sup> Mathe im Leben gGmbH, c/o Freie Universität Berlin, AG Ziegler, Arnimallee 2, 14195 Berlin, marco.traeger@gmail.com

<sup>4</sup> Mathe im Leben gGmbH, c/o Freie Universität Berlin, AG Ziegler, Arnimallee 2, 14195 Berlin, woostenfeld@mathe-im-leben.de

consecutive days in December. Teachers are encouraged to use MiA with their students and to compete in so-called class-competitions, where the entire class participates in the competition as a unit. In 2019, 45% of all users participated as part of a class in the class-competition. To access all features of the class-competition, a small fee must be paid, whereas participating as a single user is free. The virtual learning platform consists of two formats, one for children in grades 4-6 (ages 9-12) and one for children in the grades 7-9 (ages 12-15). Moreover, the MiA website is in a grey area between formal and informal learning, since partially, it is used by teachers (using, e.g., print-outs of the tasks) as a part of their curriculum. This platform implements a responsive web design based on CSS media queries and is thus usable both on desktop PCs, as well as on mobile devices. To examine the current state of usage for different devices, specifically in a learning environment, user traffic data from Google analytics was analyzed.

## 2 Background: Mobile Learning

With the progressing digitalization arriving in schools among many other areas, the demand and offer of learning options for mobile usage rises [Sh09]. With mobility and context as the main objects of analysis, theoretical frameworks of mobile learning have been published [AH07]. Some predict that mobile learning has the potential to become a new generation of distance learning [ZBD09]. In a survey of 88 educators across different countries, 78% of respondents believed that mobile learning will be essential to mainstream higher education within three to five years [ZBD09].

Using the wide range of options for task presentation and feedback, such as images, videos, or sounds available on tablets, as well as the adaptability of data-based learning environments, early adopters of these technologies hope to achieve improved learning contexts for students [FM20]. Apart from the personalization of learning environments, m-learning environments often offer parallelized social interactions between students, as well as students and teachers [Ke12]. These types of collaborations fall in line with the necessity of conversation in teaching and learning known from socio-cultural theory [Vy78].

In order to examine the educational reality of m-learning, the technology acceptance model (TAM) is often used to indicate the willingness of stakeholders to implement m-learning. The results of a systematic review by Al-Emran and colleagues indicate that the majority of these articles deal with the acceptance of m-learning among students [AMK18]. As digital natives, young students often appreciate mobile settings for educational purposes [PB11]. The two main indicators for behavioral intention of use, as suggested by the TAM model, are the perceived usefulness, as well as the perceived ease of use, which has also been confirmed in an m-learning setting [Al15]. Furthermore, it seems as if the usage of mobile technologies itself has a positive influence on the motivation of students [PGW08].

When examining the educators' attitude toward m-learning varying results were found.

One study found that younger female respondents were more engaging in m-learning resources [CC14]. Another study found that while the majority of teachers have a passive stance when it comes to the introduction of mobile technologies in classes, the likelihood of a progressive stance rises with their familiarity of the possibilities and potentials [FM20].

There are also concerns when it comes to mobile usage in educational contexts. The use of mobile devices in online social networking parallel to the completion of school work activities has been found to be disruptive [JC12], by using cognitive capacities of students. When it comes to m-learning in mathematics, approaches of self-directed and game-based learning have been implemented [Eb15]. However, collaborative learning scenarios, as required by the socio-cultural theory, are still very rare, which is likely due to the complex programming efforts and necessary environment to implement it [Eb15].

### 3 Method

The devices used for accessing the MiA webpage during the seasons of 2015 - 2019 were analyzed based on Google Analytics data. For all user sessions (meaning durations of uninterrupted usage of the platform) three types of devices were distinguished: desktop PCs, mobile phones and tablet devices.

In the first step, a list of URLs, which were of interest, was created. In the MiA environment, specific URLs are targeted towards specific users. Teachers, for example, are presented an overview of their classes achievements via a teacher-specific URL. As for the URLs leading to the tasks of the students, these are all labeled as “4-6” or “7-9”, indicating the presumed school classes, from which the respective age groups may be deduced, of the users. From internal data, it is known that the overwhelming majority of the users are students, and not teachers. Therefore, we assume the influence of teachers looking at the student tasks to be negligible and interpret the views of the task- and submission-specific URLs as student usage.

In the current analysis, URLs were split into 6 groups:

- Teacher-specific URLs, where the teachers overview classes between 4-6
- URLs leading to the sites where a task for classes 4-6 is displayed
- URLs where a task for classes 4-6 has just been submitted
- Teacher-specific URLs, where the teachers overview classes between 7-9
- URLs leading to the sites where a task for classes 7-9 is displayed
- URLs where a task for classes 7-9 has just been submitted

For each of the specified URLs, the absolute and relative numbers of sessions were determined. For each session, it was specified whether a desktop PC, a tablet or a mobile phone was used.

### 4 Results

Between the years 2015 and 2019, the virtual learning platform experienced a growth in absolute sessions per day. Furthermore, the percentage of mobile devices used for these sessions grew. Between the seasons of 2017 and 2018, the mobile usage overtook the desktop usage of the website as the most frequently used medium for MiA, as visible in Figure 1.

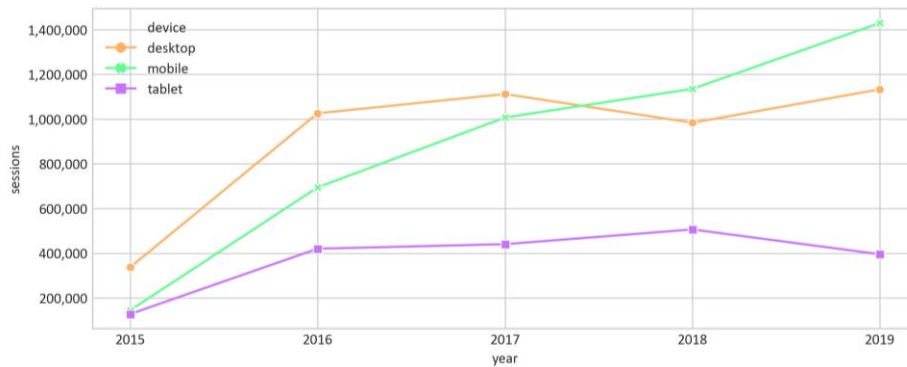


Fig. 1: Number of sessions per year, between the years 2015 and 2019, split by different devices.

It turns out that device usage differs between students and teachers (Figure 2, left). In 2019, teachers preferred desktop PCs over mobile devices for accessing MiA, students used both desktop PCs and mobile devices in a balanced way, with older students showing a slight preference for mobile devices instead over tablets.

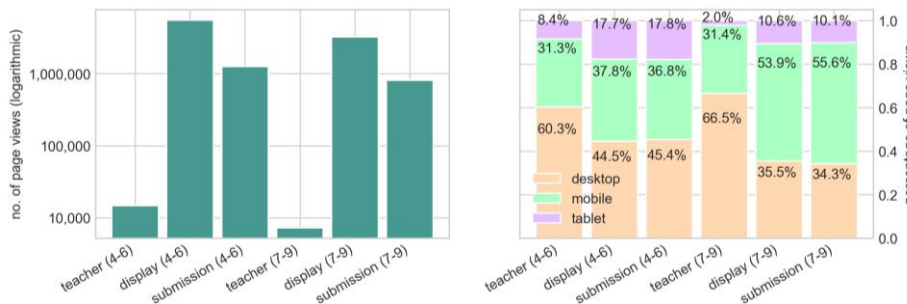


Fig. 2: Left: Average number of sessions (per task) on display or submission sites of the tasks for classes 4-6 and 7-9, separate for teachers. Right: relative numbers of sessions on different devices by teachers and students of classes 4-6 and 7-9.

As can be seen on the right in Figure 2, mobile devices were preferred for displaying tasks. When submitting the tasks, this relation shifted towards desktop PCs for the younger students (grades 4-6). While the usage of desktop PCs also increased, the students in grades 7-9 largely submitted the tasks via mobile devices.

## 5 Conclusion

All in all, we have confirmed that the trend of website usage on mobile devices rises, also in an educational context. Specifically, we have detected a rise of mobile usage over the past five years, which overtook the usage of desktop PCs.

Secondly, we have seen that teachers present a different media behavior than students. The prevalent use of mobile devices for displaying and submitting e-learning tasks was especially visible for students of grades 7-9. There were no notable differences in the preferred devices between the display and the submission of the tasks.

We conclude that, especially for the age group of 12- to 15-year-olds, mobile devices have become a predominant medium for e-learning content. The practical implications of these findings are that the differing needs for optimal usage of teachers and students must be involved in the design process of e-learning materials, for example in the forms of responsive or adaptive website design. Furthermore, learning content must also be adapted for optimal mobile usage in the case of 12- to 15-year-olds.


As an outlook for further research, one could further distinguish between the user groups by applying attribution modeling and tag management within Google Analytics. This would yield more insight about the contexts and further activities of the learners.

## Bibliography

- [AES16] Al-Emran, M.; Elsherif, H. M.; Shaalan, K.: Investigating attitudes towards the use of mobile learning in higher education. In *Computers in Human Behavior*, 2016, 56; S. 93–102.
- [AMK18] Al-Emran, M.; Mezhuyev, V.; Kamaludin, A.: Technology Acceptance Model in M-learning context: A systematic review. In *Computers & Education*, 2018, 125; S. 389–412.
- [AI15] Althunibat, A.: Determining the factors influencing students' intention to use m-learning in Jordan higher education. In *Computers in Human Behavior*, 2015, 52; S. 65–71.
- [AH07] Andrews, R.; Haythornthwaite, C.: *The Sage handbook of e-learning research*. Sage, 2007.
- [CC14] Mark Anthony Camilleri; Adriana Caterina Camilleri: *Measuring the educators' behavioural intention, perceived use and ease of use of mobile technologies*, 2017.

- 
- [CP17] Cazañas, A.; Parra, E.: Strategies for mobile web design. In *Enfoque UTE*, 2017, 7; S. 344–357.
- [Eb15] Ebner, M.: Mobile applications for math education-how should they be done. In *Mobile learning and mathematics. Foundations, design, and case studies*, 2015; S. 20–32.
- [FM20] Forkosh-Baruch, A.; Meishar-Tal, H.: Proactive, Preventive or Indifference?: Reaction Modes of Faculty towards Use of Personal Mobile Devices in Courses: Mobile Devices in Education: Breakthroughs in Research and Practice. IGI Global, 2020; S. 298–311.
- [JC12] Junco, R.; Cotten, S. R.: No A 4 U: The relationship between multitasking and academic performance. In *Computers & Education*, 2012, 59; S. 505–514.
- [Ke12] Kearney, M. et al.: Viewing mobile learning from a pedagogical perspective. In *Alt-J-Research In Learning Technology*, 2012, 20.
- [Mu14] Mullins, C.: Responsive, mobile app, mobile first: untangling the UX design web in practical experience: Proceedings of the 33rd Annual International Conference on the Design of Communication, 2015; S. 1–6.
- [PB11] Pollara, P.; Broussard, K. K.: Student perceptions of mobile learning: A review of current research: Society for information technology & teacher education international conference, 2011; S. 1643–1650.
- [PZ15] Peng, W.; Zhou, Y.: The design and research of responsive web supporting mobile learning devices: 2015 International Symposium on Educational Technology (ISET), 2015; S. 163–167.
- [RGW08] Rau, P.-L. P.; Gao, Q.; Wu, L.-M.: Using mobile communication technology in high school education: Motivation, pressure, and learning performance. In *Computers & Education*, 2008, 50; S. 1–22.
- [Sh09] Sharples, M. et al.: *Mobile learning: Technology-enhanced learning*. Springer, 2009; S. 233–249.
- [St19] StatCounter. *Desktop vs Mobile vs. Tablet Market Share Worldwide*, 2019.
- [Vy78] Vygotsky, L.: Interaction between learning and development. In *Readings on the development of children*, 1978, 23; S. 34–41.
- [Wa15] Walsh, K.: Mobile learning in medical education. In *Ethiopian Journal of Health Sciences*, 2015, 25; S. 363–366.
- [ZBD09] Zawacki-Richter, O.; Brown, T.; Delpont, R.: Mobile Learning: From single project status into the mainstream? In *European Journal of Open, Distance and E-Learning*, 2009, 12.

## Digitalisierung von Analyse- und Auswertungsstrukturen im Kontext schulischer Wettbewerbsszenarien

Florian Funke <sup>1</sup> und Sven Hofmann<sup>2</sup>

**Abstract:** Im Rahmen des Mathematik-Wettbewerbs Run For Numbers nehmen Schülerinnen und Schüler an Schulen in den Bundesländern Sachsen, Thüringen und Brandenburg halbjährlich teil, um ihre mathematischen Fähigkeiten und Fertigkeiten zu testen. Mit Hilfe des siebenminütigen Speed-Tests soll ein möglichst umfassendes Bild des aktuellen Lernstands erfasst werden. Im folgenden Artikel werden die umgesetzten Maßnahmen zur Digitalisierung der Auswertung und Analyse der Wettbewerbsdaten sowie dabei eingesetzte Kenngrößen, Maße und Normen beschrieben. Es wird ebenso dargestellt, inwieweit die Verständlichkeit der Werte für Lernende und Lehrende im schulischen Kontext sichergestellt wird und welche Darstellungsformen und Formulierungen eingesetzt werden.

**Keywords:** Assessment, Digitalisierung, Datenanalyse, Schule, Wettbewerb, Feedback, automatisierte Rückmeldung

### 1 Einleitung

Um fachübergreifende mathematische Schlüsselkompetenzen zu trainieren und eine Rückmeldung zum aktuellen Lernstand zu geben, wurde der Wettbewerb Run For Numbers entwickelt. Damit dieser eine hohe Passung in die schulischen Rahmenbedingungen aufweist, wurde er als siebenminütiger Speed-Test konzipiert. Eine Einbettung des Wettbewerbs als Teil einer Unterrichtseinheit ist damit sehr gut möglich. Zur Qualitätssicherung und iterativen Verbesserung werden Beurteilung und Bewertung der Lösungen, sowie die Auswertung und Analyse zentral von der Wettbewerbsorganisation durchgeführt. Für eine Anpassung an individuelle Voraussetzungen wird die Durchführung digital und papierbasiert angeboten [Fu19a]. Durch die laufend veränderlichen Rahmenbedingungen für die Durchführung, beispielsweise in der Ausstattung der Schulen, wurden Möglichkeiten untersucht, die Analyse und Auswertung der Wettbewerbsdaten unabhängig von der Art und Weise der Durchführung zu gestalten. Damit wird sichergestellt, dass bei Veränderungen der Rahmenbedingungen des Wettbewerbs keine umfassende Rekonstruktion der Analysestrukturen erfolgen muss. Im Folgenden werden die implementierten Funktionen und die verwendeten Methoden erläutert sowie die daraus generierten Feedbacks für Lernende und Lehrende beschrieben.

---

<sup>1</sup> Universität Leipzig, Didaktik der Informatik, Augustusplatz 10, 04109 Leipzig,  
florian.funke@informatik.uni-leipzig.de,  <https://orcid.org/0000-0003-4043-894X>

<sup>2</sup> Universität Leipzig, Didaktik der Informatik, Augustusplatz 10, 04109 Leipzig,  
sven.hofmann@informatik.uni-leipzig.de

## 2 Datenerhebung und Datenverarbeitung

Die Konzeption der Analysewerkzeuge sollte mehrere veränderliche Rahmenbedingungen des Wettbewerbs berücksichtigen. Zum einen findet der Wettbewerb nach wie vor auf Grund unterschiedlicher Ausstattungen und Voraussetzungen an Schulen papierbasiert und digital statt. Parallel dazu steigen die Teilnehmerzahlen kontinuierlich, auf zuletzt über 5000 teilnehmende Schülerinnen und Schüler, an [Fu19a]. Andererseits sind neue Wege der Rückmeldung auf verschiedenen Ebenen eingebunden worden. So sind nun Möglichkeiten des individualisierten Feedbacks für Teilnehmende und Lehrende etabliert. Mit Hilfe der neu geschaffenen Strukturen wird darüber hinaus die Qualitätssicherung des Wettbewerbs vorgenommen. Der Ablauf der Datenerhebung ist in Abb. 1 dargestellt.

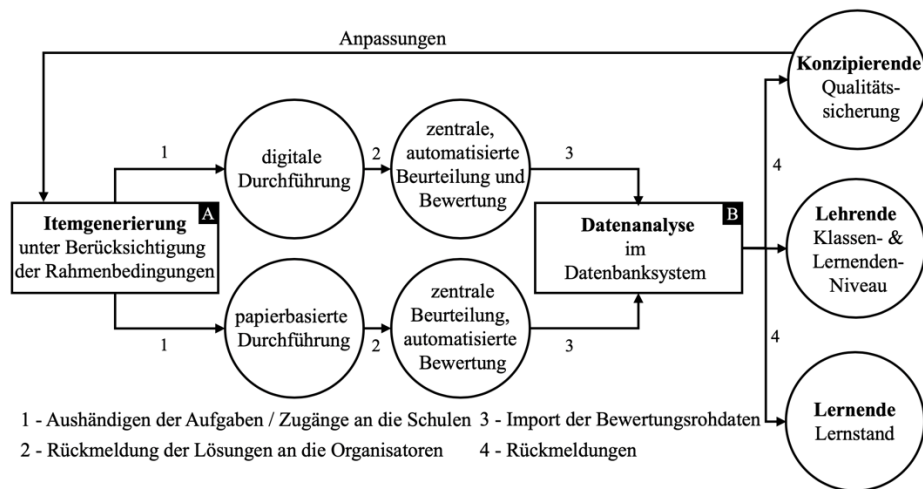


Abb. 1: Ablauf der Datenerhebung im Rahmen des Run For Numbers

## 3 Datenanalyse

Um die auf verschiedenen Wegen erhobenen Daten zentral analysieren zu können, wurde ein Datenbanksystem entwickelt, in dem sich die Rohdaten mittels Import-Funktion, aber auch händisch einpflegen lassen. Mit Hilfe von Einweg-Hash-Funktionen wird ermöglicht, dass die Daten eines Teilnehmers ihm nicht mehr zuzuordnen sind. Da die Berechnung aber über die Wettbewerbe hinweg aus den gleichen Metadaten erfolgt, ist weiterhin die Zuordnung von verschiedenen Teilnahmen am Wettbewerb zu einem Teilnehmer möglich. Veränderungen des individuellen Lernstands sind damit in zukünftig entwickelten Darstellungsformen verwendbar und können rückgemeldet werden. Die innerhalb eines Wettbewerbs erhobenen Daten werden dabei auf Ebene der Items, der Teilnehmer und der Klasse untersucht.



### 3.1 Itemanalyse

Die Itemanalyse leistet einen Beitrag hinsichtlich der Untersuchung von Leistungen in Korrelation mit der Gesamtstichprobe sowie bei der Erhöhung der Validität und damit in der Qualitätssicherung. Die hohe Validität eines Itemsatzes impliziert, dass Personen mit starker Merkmalsausprägung die angegebenen Aufgaben häufiger richtig beantworten, als Personen, die eine schwache Merkmalsausprägung haben [Li89]. Die Itemanalyse kann dabei zur kontinuierlichen Erhöhung der Validität über mehrere Iterationen beitragen. Die Validität bestimmt sich über die Objektivität, Reliabilität, Schwierigkeit und Trennschärfe [Li89]. Bei der Bestimmung dieser Größen erfolgt eine Orientierung an der klassischen Testtheorie, da sich die untersuchte mathematische Kompetenz aus verschiedenen Dimensionen zusammensetzt. Eine Modellierung konformer Items im Rahmen der Item-Response-Theorie ist mit dieser Annahme nicht möglich [Ro04].

Die Prüfung der Objektivität erfolgt in den Dimensionen Durchführung, Auswertung und Interpretation. Die Sicherstellung erfolgt mit Hilfe von standardisierten Abläufen und Formulierungen, die gleichwertige Bedingungen für alle Wettbewerbsteilnehmer garantieren sollen. Insbesondere im Rahmen der zentralen, teilautomatisierten Beurteilung und Bewertung kann die Objektivität sehr gut sichergestellt werden. Dabei hat die Formalisierung der Aufgabenstellungen hohe Relevanz, um mögliche Alternativlösungen zu eliminieren und damit die automatisierte Beurteilung und Bewertung zu ermöglichen. Auf Basis der stattfindenden Datenanalyse wurde die Objektivität kontinuierlich von Wettbewerb zu Wettbewerb verbessert. Dabei erfolgt die Formalisierung der verwendeten Items in mehreren Schleifen. Der Ablauf einer Iteration ist in Abb. 1 dargestellt. Die im Folgenden beschriebenen Kenngrößen der Itemanalyse bieten wertvolle Indikatoren, um problematische Items zu identifizieren, anzupassen oder zu eliminieren.

Für die Bestimmung der Schwierigkeit eines Items wird der Schwierigkeitsindex herangezogen. Dieser gibt den Anteil richtiger Antworten eines Items an [Li89]. Der Schwierigkeitsindex findet unter anderem Anwendung in der Analyse einer Klasse im Vergleich zur Gesamtstichprobe. Lehrende können schnell vergleichen, wie ein Item in der Klasse im Vergleich zum gesamten Teilnehmerfeld beantwortet wurde. Daraus können Rückschlüsse über den Lernstand der Klasse im untersuchten mathematischen Teilgebiet gezogen und Aufgabentypen, die zu leicht oder zu schwer sind, identifiziert werden.

Die Trennschärfe beschreibt die Korrelation der Beantwortung eines Items mit der Gesamtpunktzahl des Teilnehmenden [Fi04]. Im Rahmen der Test- und Itemkonstruktion wird auf Basis vorheriger Wettbewerbe versucht, Items mit einer Trennschärfe im Bereich .5 bis .9 zu generieren. Dies erfolgt, da Items mit zu hoher Trennschärfe (>.9) wenig neue Informationen über die Merkmalsausprägung liefern und Items mit geringer Trennschärfe (<.5) andere Eigenschaften, als die Anvisierten bestimmen [Ro04]. Damit wird eine gute Differenzierung zwischen den Teilnehmenden ermöglicht. Dies gelingt noch nicht umfänglich. Im Herbst 2019 waren lediglich jedes Zehnte der Items zwischen .5 und .9, aber zwei von drei Items in der Umgebung des anvisierten Bereichs (.4 bis .95) [Fu19b].

Für die Bestimmung der Reliabilität wird das Kuder-Richardson-Alpha verwendet. Dieses gibt an, inwieweit die Items innerhalb einer Testung Gleiches messen und die Äquivalenz sowie Stabilität von Messungen gewährleistet ist [Li89]. Mit Hilfe des berechneten Maßes ist es dementsprechend möglich die Qualität des Itemsets einzuschätzen. Dabei werden Werte zwischen .65 und .95 für Testungen als gut angenommen [GM02]. Im Herbst 2019 wurden Werte zwischen .76 und .84 erreicht [Fu19b].

### **3.2 Teilnehmeranalyse**

Im Rahmen der Analyse der Teilnehmer soll eine zweistufige Verortung der Leistung eines Teilnehmers im gesamten Teilnehmerfeld vollzogen werden. Es kann in der Auswertung ein Vergleich mit zurückliegenden Leistungen (intraindividuelle Norm), und mit Leistungen anderer Gruppen im aktuellen Wettbewerb (interindividuelle Norm) erfolgen. Vergleichsgruppen sind hierbei Klassen, Schulen und Schulformen.

Im ersten Schritt wird der Prozentrang jedes Teilnehmenden ermittelt. Dieser gibt an, wie viel Prozent aller Teilnehmer die gleiche oder eine geringere Punktzahl erreicht haben [Li89]. Der Prozentrang wird daher zur Verortung der individuellen Leistung eines Teilnehmenden in der Stichprobe eingesetzt. Der entstandene Wert wird an Lernende und Lehrende gleichermaßen rückgemeldet. Mit den Werten der Lernenden können der aktuelle Lernstand sowie Tendenzen in der Entwicklung abgelesen werden. Der Prozentrang ist mit dem Charakter einer Stichprobe versehen, da er nicht normiert ist, auch wenn die Verteilung der Teilnehmer einer Normalverteilung in der Regel ähnelt [Fu19b].

Zur Normierung wird aus diesem Grund der T-Wert als weiteres Maß aus dem Prozentrang berechnet. Die T-Wert-Skala eignet sich in diesem Fall, da sie immer dann Anwendung findet, wenn die Messwerte keine Normalverteilung liefern, aber eine solche für das untersuchte Merkmal angenommen wird [HL11]. Damit wird mit Hilfe des T-Werts, ähnlich dem Prozentrang, die Position eines Teilnehmers innerhalb der gesamten Messung widerspiegelt, gleichzeitig aber der Stichprobencharakter eliminiert.

Aus den beschriebenen Kenngrößen werden über die individuelle Rückmeldung hinaus Klassenwerte bestimmt. Dabei eignet sich der Mittelwert der T-Werte besonders, um die Klasse im Vergleich zur Gesamtstichprobe zu betrachten. Mit Hilfe des Prozentrangs wird dargestellt, wie viel Prozent der Teilnehmer einer Klasse im Schnitt überboten wurden.

## **4 Aufbereitung für Teilnehmende und Lehrende**

Für Empfänger, die im Umgang mit den berechneten Werten und ihren Bedeutungen keine Erfahrung besitzen, müssen diese interpretierbar gemacht werden. Auch stellt sich die Frage, welche Werte in welcher Form rückgemeldet werden. Die Auswertung erfolgt dabei in zwei Parallelen für die Lernenden und Lehrenden (vgl. Abb. 2).

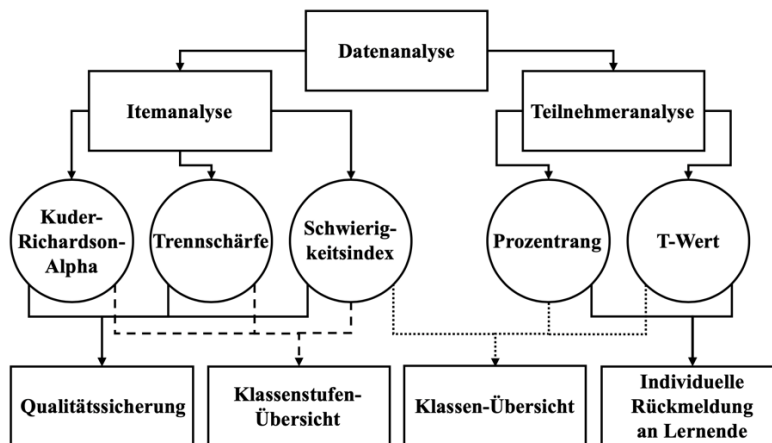


Abb. 2: Berechnung und Verwendung der Kenngrößen

Für die Teilnehmenden und ihre Orientierung an interindividuellen Normen stellen der Prozentrang und der T-Wert zentrale Rückmeldungen dar. Dabei lässt sich der Prozentrang gut in eine leichter interpretierbare Kenngröße überführen. Den Schülerinnen und Schülern wird die Aussage: „Du gehörst zu den besten x% deiner Klassenstufe.“ rückgemeldet. Diese ist für sie besser verständlich und gut interpretierbar.

Aus dem T-Wert lässt sich ein Ausschnitt des Teilnehmerfeldes ermitteln, in dem sich die Teilnehmenden bewegen. Anhand festgelegter Bereiche, in denen sich der T-Wert eines Teilnehmers befindet, werden verschiedene sprachliche Rückmeldungen gegeben. Für Teilnehmende mit einem T-Wert im Bereich 60 bis 70 wird beispielsweise „Du hast herausragende Leistungen erbracht“ rückgemeldet.

Für die Lehrenden soll ein umfassender statistischer Überblick zum Wettbewerb, zur Klasse und für die einzelnen Lernenden möglich sein. Daher werden zunächst die Daten der Itemanalyse und statistische Daten, wie die im Mittel erreichten Punkte und deren Verteilung in einer Klassenstufenübersicht aufbereitet. Hierbei werden Items mit hohem und niedrigem Schwierigkeitsindex farblich hervorgehoben, so dass beispielsweise besonders schwere Aufgaben in der Klasse zur Diskussion herausgehoben werden können.

Darüber hinaus erhalten die Lehrenden eine Übersicht, wie die Klasse Aufgaben im Mittel beantwortet hat. Hierbei werden analog Abweichungen von der Gesamtbeantwortung und damit dem Schwierigkeitsindex hervorgehoben. Hier werden klassenspezifische Exzellenz- und Potenzialgebiete dargestellt. Abschließend wird eine Übersicht mit den Rohwerten, T-Werten und Prozenträngen sowie den zugehörigen generierten sprachlichen Formulierungen zur Verfügung gestellt, so dass auch die individuellen Werte der Lernenden von den Lehrenden untersucht werden können. Die Interpretierbarkeit der Ergebnisse wird dabei durch eine Kurzhandreichung sichergestellt, die die Hintergründe der einzelnen Werte darstellt. Alle beschriebenen Übersichten werden automatisch generiert und in eine gut verteilbare Form, beispielsweise Serienbriefe, gebracht.

## 5 Fazit und Ausblick

Die für den Wettbewerb Run For Numbers umgesetzten zentralisierten, digitalen Analysestrukturen konnten etabliert werden. Sie haben sich in den Wettbewerben als hilfreich bei der Generierung von individualisiertem Feedback in verschiedenen Dimensionen gezeigt. Durch sie konnten bei kontinuierlich steigenden Teilnehmerzahlen umfassendere und hilfreiche Auswertungen für Lernende und Lehrende generiert werden. Die Rückmeldungen der teilnehmenden Schulen zeigen auf, dass durch die regelmäßige Durchführung, und dass damit einhergehende Feedback an die Lernenden, gezielt individuelle Förderungen angestoßen werden können. Die eingesetzten Erläuterungen und Verbalisierungen erzeugen eine Verständlichkeit der statistischen Werte und machen sie erst dadurch insbesondere für die Lernenden nutzbar.

Im Rahmen der Weiterentwicklung sollen gezielt auch wettbewerbsübergreifende Daten ausgewertet und in einer intuitiven Form dargestellt werden, damit die Lernenden und Lehrenden nicht nur eine Rückmeldung zum aktuellen Lernstand, sondern auch zur Entwicklung im Vergleich zu vorangegangenen Wettbewerben erhalten. Hierbei können die bereits verwendeten statistischen Werte eingebunden und visualisiert werden. Darüber hinaus werden Konzepte zur Abbildung der Erfüllung kriterialer Normen entwickelt und erprobt.

## Literaturverzeichnis

- [Fi04] Fisseni, H.: Lehrbuch der psychologischen Diagnostik. Göttingen: Hogrefe, S. 36ff, 2004.
- [Fu19a] Funke, F.: Entwicklung und Evaluation der iterativen Digitalisierung kompetitiver Assessment-Szenarien dargestellt am Beispiel des Wettbewerbs Run For Numbers, 2019.
- [Fu19b] Funke, F.: Ergebnisse Herbst 2019. <https://runfornumbers.de/ergebnisse-herbst-2019/>, Stand: 26.03.2020
- [GM02] George, D.; Mallery, P.: SPSS for Windows Step by Step. Needham Heights, MA: Pearson Higher Education, S.231, 2002.
- [HL11] Hesse, I.; Latzko, B.: Diagnostik für Lehrkräfte. Opladen: Budrich, S. 75, 77, 2011.
- [Li89] Lienert, G.: Testaufbau und Testanalyse. München: Psychologie-Verl.-Union, S. 38, 40, 70, 226, 331ff., 1989.
- [RK39] Richardson, M. W.; Kuder, G. F.: The Calculations of Test Reliability Coefficients Based on the Method of Rational Equivalence. J. Educ. Psychol, S. 30, 681, 1939.
- [Ro04] Rost, J.: Lehrbuch Testtheorie - Testkonstruktion. Bern: Huber. S. 218, 2004.

## Fertigkeiten erkennen und Interesse für Informatik wecken: Der Informatiktest von MINTFIT Hamburg

Helena Barbas<sup>1</sup>, Esther Bender<sup>2</sup>, Fabian Hamann<sup>3</sup>, Daniel Sitzmann<sup>4</sup>, Marcus Soll<sup>5</sup>

**Abstract:** Das Verbundprojekt MINTFIT der Hamburger Hochschulen entwickelt seit 2014 kostenfrei und anonym nutzbare Online-Selbsteinschätzungstests und E-Learning-Angebote für Schüler\*innen und Studieninteressierte. Der Themenumfang war bisher beschränkt auf die Fächer Mathematik, Physik und Chemie; Kenntnisse wurden auf Basis des Schulwissens überprüft (Nutzung bundeseinheitlicher Mindestanforderungskataloge, falls vorhanden). Im Zuge der Erweiterung um den Bereich Informatik bestand die Herausforderung, dass es kein bundesweites Pflichtschulfach Informatik in Deutschland gibt und somit neue Strategien für die Auswahl der Themen (fehlende bzw. stark heterogene Bildungspläne) und die Fragenerstellung (keine Abfrage von Schulwissen möglich) notwendig wurden. Deshalb wurde ein Test entwickelt, der mit einleitenden Erklärungen und aufeinander aufbauenden Teilaufgaben das grundsätzliche Verständnis von für Informatik relevanten Themen prüft. Dieser Beitrag zeichnet die Entwicklung des MINTFIT-Selbsteinschätzungstests für Informatik nach und beschreibt einerseits, wie die Themengebiete u.a. anhand einer bundesweiten Umfrage unter Lehrenden und durch Vergleich vorhandener Lehrpläne gefunden und gewichtet wurden. Andererseits werden Testaufbau, Entwicklung und Auswahl von Fragen betrachtet sowie erste Evaluationsergebnisse anhand von Erprobungen in Schulen und des in einer Beta-Version veröffentlichten Tests vorgestellt.

**Keywords:** Self-Assessment in Informatik, Übergang Schule-Hochschule, E-Learning, MINTFIT.

### 1 Einleitung – Was ist MINTFIT?

Die Förderung der MINT-Ausbildungsberufe und -Studienfächer und damit auch Maßnahmen zur Senkung der durchschnittlich sehr hohen Abbruchquoten in den MINT-Studiengängen (je nach Fach um die 30%, siehe [He17]) sind Ziele aktueller Bildungspolitik. Hamburger Hochschulen setzen seit 2014 am Übergang Schule-Hochschule mit dem Verbundprojekt MINTFIT Hamburg an: Anhand webbasierter Selbsteinschätzungstests unter [www.mintfit.hamburg](http://www.mintfit.hamburg) können Schüler\*innen und Studieninteressierte ihre MINT-Kenntnisse selbstständig dahingehend überprüfen, ob

---

<sup>1</sup> HafenCity Universität Hamburg, MINTFIT, Geomathematik, -informatik & Physik, Überseeallee 16, 20457 Hamburg, [helena.barbas@hcu-hamburg.de](mailto:helena.barbas@hcu-hamburg.de)

<sup>2</sup> Hochschule für Angewandte Wissenschaften Hamburg, MINTFIT, Alexanderstraße 1, 20099 Hamburg, [esther.bender@haw-hamburg.de](mailto:esther.bender@haw-hamburg.de)

<sup>3</sup> Technische Universität Hamburg, MINTFIT, Institut für Mathematik, Am Schwarzenberg-Campus 3, 21073 Hamburg, [fabian.hamann@tuhh.de](mailto:fabian.hamann@tuhh.de)

<sup>4</sup> Universität Hamburg, MINTFIT, Max-Brauer-Allee 60, 22767 Hamburg, [daniel.sitzmann@uni-hamburg.de](mailto:daniel.sitzmann@uni-hamburg.de)

<sup>5</sup> Universität Hamburg, MINTFIT, Max-Brauer-Allee 60, 22767 Hamburg, [marcus.soll@uni-hamburg.de](mailto:marcus.soll@uni-hamburg.de)

diese den Anforderungen an ein MINT-Studium an einer deutschen Hochschule entsprechen. Sofort nach Testabgabe erhalten Teilnehmer\*innen eine automatische Auswertung inklusive Musterlösungen. Zudem werden individuelle Lernempfehlungen generiert, die auf relevante Kapitel in den zugehörigen E-Learning-Angeboten verweisen und mit denen so die identifizierten Wissenslücken geschlossen werden können. Das Niveau der MINTFIT-Tests ist auf Abiturient\*innen ausgerichtet (siehe [BS18], [Me19], [MS19]). Es wird Schulwissen der Sekundarstufe I und II abgefragt, die Themen wurden anhand bundeseinheitlicher Mindestanforderungskataloge (wie COSH für Mathematik, siehe [co14]) oder Vergleich der Bildungspläne der Länder ausgewählt.

Derzeit deckt das MINTFIT-Angebot die Fächer Mathematik und Physik ab. Seit 2018 werden Tests (Beta-Versionen sind bereits verfügbar) und E-Learning-Angebote in Chemie und Informatik entwickelt, was Ende 2020 abgeschlossen sein soll.

MINTFIT Hamburg ist auch aufgrund zahlreicher Kooperationen mit über 30.000 Testteilnahmen jährlich (2019) heute einer der am meisten genutzten Selbsteinschätzungstests in Deutschland.

## **2 Struktur des MINTFIT Informatiktests**

Der MINTFIT Informatiktest ist unter [www.mintfit.hamburg/informatik-test](http://www.mintfit.hamburg/informatik-test) (derzeit als Beta-Version) erreichbar. Der Test kann anonym oder mit registriertem Profil abgelegt werden. Die Bearbeitungszeit ist auf 60 Minuten angelegt und kann individuell unter- oder überschritten werden; als Hilfsmittel werden nur Stift und Papier benötigt.

Der Informatiktest gliedert sich in fünf Themenblöcke (Logik, Programmieren, Algorithmen, Formale Sprachen, Rundblick Informatik, siehe Abschnitt 3) zu je vier Fragen, die aufeinander aufbauen. Zu diesem Zweck wird eine Frage erst nach Beantwortung der vorigen angezeigt; Musterlösungen können direkt nach Bearbeitung einer Frage eingeblendet werden. Nach Testabgabe erfolgt eine automatisierte Auswertung mit textlichem und optischem Feedback in Form einer Teilnehmermedaille, die für gute bis sehr gute Leistungen auch in Bronze, Silber oder Gold angezeigt wird.

Themenauswahl und Testkonzept des Informatiktests unterscheiden sich von den Tests anderer Fachgebiete der MINTFIT-Plattform. Da es bislang kein bundesweites Pflichtschulfach Informatik gibt, existieren weder einheitliche Bildungspläne der Länder noch Mindestanforderungskataloge; Bildungsstandards der Gesellschaft für Informatik (siehe [Bi19]) sind nicht bindend für den Unterricht. Folglich verfügen Studieninteressierte bezüglich Umfang und Niveau über stark heterogenes Vorwissen, und es darf zur Bearbeitung des Informatiktests kein Vorwissen notwendig sein. Stattdessen werden hauptsächlich Soft Skills wie logisches Denkvermögen, konzentriertes Arbeiten, Problemlösekompetenz, Auge für Details sowie Englischkenntnisse in Fragen zu den fünf genannten fachlichen Themengebieten überprüft, die einen ersten Eindruck von Informatik-Inhalten geben.

### 3 Entwicklung des MINTFIT Informatiktests

Die Entwicklung des MINTFIT Informatiktests wurde im Herbst 2018 begonnen. Es wurde zunächst eine Themen-Vorauswahl auf Grundlage der Informatik-Bildungspläne der Länder, der Hamburger Hochschul-Curricula und bestehender Informatiktests (z.B. [Te20]) getroffen (siehe [Si19]). Diese wurde auf Basis einer deutschlandweiten Umfrage unter Informatik-Hochschuldozent\*innen angepasst und validiert (siehe [Do20]).

#### 3.1 Endgültige Themenauswahl auf Basis der Dozent\*innenumfrage

Für die Anfang bis Mitte 2019 durchgeführte Dozent\*innenumfrage wurden deutschlandweit ca. 6000 Informatik-Dozent\*innen aller Statusgruppen angeschrieben. Die aus den 588 Rückmeldungen gewonnenen Meinungen und Gewichtungen wurden mit eigenen Themenvorstellungen und Recherchen verglichen. Als endgültige Themenauswahl kristallisierte sich folgende heraus:

- **Logik** ist in der Studieneingangsphase relevant; logisches Denkvermögen lässt sich zudem mittels Logik-Rätseln gut abprüfen. Die Mehrheit der befragten Dozent\*innen hält dieses Themengebiet für den Test für sinnvoll.
- **Programmieren** ist Bestandteil vieler Studiengänge mit Informatik-Inhalten. Programmierkenntnisse können nicht vorausgesetzt werden, Teilnehmer\*innen sollen in einigen Aufgaben aber einfache Programme verstehen. Auch dieses Themengebiet hält die Mehrheit der Dozent\*innen für sinnvoll.
- **Algorithmen** wurden wie die beiden ersten Themengebiete von der Mehrheit der Dozent\*innen als sinnvoll für den Test eingestuft. Mithilfe dieses Themengebiets können Problemlösestrategien besonders gut getestet werden.
- **Formale Sprachen:** Dieses Themengebiet wird nicht von einer Mehrheit der Dozent\*innen als für den Test sinnvoll erachtet, dafür schätzt die Mehrheit es bei Studierenden aber als problembehaftet ein. Wenn Studieninteressierte im Test mit dem Themengebiet bereits in Berührung gekommen sind, führt es im Studium womöglich zu weniger Problemen. Zusätzlich wird so vermittelt, dass es in der Informatik auch vermehrt theoretische und abstrakte Aspekte gibt, was vielen Studieninteressierten unter Umständen nicht bewusst ist (siehe [He10]).
- **Rundblick Informatik:** Studieninteressierten soll hier mit nichttrivialen Fragen zu verschiedenen Themen gezeigt werden, dass Informatik nicht nur aus Programmierung besteht. Dieses Vorurteil (siehe [MW06]) abzubauen war ein mehrfach geäußelter Wunsch in den Freitextfeldern der Dozent\*innenumfrage.

### **3.2 Testkonzept**

Wie in Teil 2 erwähnt, soll der MINTFIT Informatiktest kein Vorwissen abfragen. Alle für den Test benötigten Grundlagen müssen daher dort erklärt werden. Damit trotzdem komplexere Fragestellungen möglich sind, bauen die Fragen eines Themengebiets (ausgenommen „Rundblick Informatik“) aufeinander auf. So können Grundlagen, die in den Fragen erklärt werden, auch in den folgenden Aufgaben verwendet werden. Der Test nutzt dabei Mechanismen, die verhindern, dass Nutzer\*innen Folgefehler begehen oder Fragen bearbeiten, zu der wichtige Informationen fehlen. Erstens können Nutzer\*innen nach Bearbeitung einer Frage die Musterlösung aufdecken, um sofort mögliche fehlerhafte Vorstellungen zu erkennen und Folgefehler zu vermeiden. Zweitens kann eine weitere Frage erst dann angezeigt werden, wenn die vorige beantwortet wurde. Darüber hinaus werden die aufeinander aufbauenden Fragen genutzt, um verschiedene Schwierigkeitsgrade abzubilden. Die erste Frage soll dabei als Motivation ein leichter, spielerischer Einstieg ins Themengebiet sein. Der von Aufgabe zu Aufgabe steigende Schwierigkeitsgrad spricht so auch Teilnehmende mit verschiedenen Wissensständen an.

### **3.3 Fragenerprobung und -auswahl**

Im Zuge der Entwicklung des Informatiktests wurden ca. 160 Fragen erstellt, die von Fachwissenschaftler\*innen überprüft und iterativ verbessert wurden. Mit einer Auswahl von ca. 40 Fragen aus diesem Pool wurden Tests zusammengestellt und von Schüler\*innen bei Schulbesuchen erprobt. Mittels Feedbackbögen, Beobachtungen und Gesprächen während der Durchführung sowie Testauswertungen wurde ein aus 20 Fragen bestehender finaler Test erstellt. Dieser wurde erneut mit Schüler\*innen erprobt, bevor er schließlich als Beta-Version des MINTFIT Informatiktests veröffentlicht wurde.

Beispiele für die Fragen des MINTFIT Informatiktests sind z.B. im Bereich Logik das Bewerten des Wahrheitsgehalts von Aussagen basierend auf verschiedenen Prämissen; im Bereich Programmieren wird thematisch ein Haushaltsroboter genutzt, an dem verschiedene Anweisungen und Konstruktionen von Programmiersprachen abgefragt werden. Der Bereich Rundblick Informatik ist atypisch, da er zu verschiedenen Themen (Binärzahlen, Datenbanken und Netzwerke, Hardware) jeweils eine Frage enthält; zudem werden hier Englischkenntnisse sowie Lesekompetenz in Englisch abgeprüft.

### **3.4 Testvarianten und ihre Einsatzzwecke**

Für unterschiedliche Zwecke wurden Kurzversionen des Informatiktests entwickelt. Eine Variante kann in einer Doppelunterrichtsstunde absolviert werden und lässt Zeit für ausführliches Besprechen der Lösungen. Für Studieninformationstage wurde eine Teaser-Variante des Tests mit wenigen, kurzen Fragen erstellt, die Interessierten einen ersten Eindruck vermitteln und Interesse an der Standardversion des Tests wecken soll.



## **4 Bisherige Erfahrungen mit dem MINTFIT Informatiktest**

Ende November 2019 ging die Beta-Version des MINTFIT Informatiktests online. Auch nach der Veröffentlichung wurde noch eine Testreihe an Schulen durchgeführt, um die Qualität des Tests weiter zu verbessern. Zudem wurden die Ergebnisse der mittlerweile ca. 1000 durchgeführten Testversuche (Stand Juni 2020) ausgewertet. Dabei wurde zum Beispiel untersucht, welche Fragen eventuell Hürden darstellen, weil es nach der Bearbeitung verhältnismäßig viele Testabbrüche gab. So wurde der Schwierigkeitsgrad einiger Fragen angepasst; das Anzeigen der Musterlösungen während des Testversuchs wurde auf optional umgestellt, um den Arbeitsfluss während des Tests zu verbessern.

Nach Testabgabe werden Teilnehmer\*innen gebeten, einen Feedbackbogen zum Informatiktest auszufüllen - dieser beinhaltet u.a. Fragen zum Schwierigkeitsgrad und ob der Test als interessant und die Erklärungen als verständlich eingestuft werden. Teilnehmer\*innen der Befragung schätzten ihr Informatik-Vorwissen als „mittel“ bis „wenig“ ein, bestätigten, dass kein Vorwissen zum Ablegen des Tests benötigt wurde, stuften Musterlösungen als hilfreich und den Schwierigkeitsgrad der Fragen als „mittel“ bis „schwierig“ ein, hatten größtenteils Spaß bei der Bearbeitung und fanden den Test interessant. Im Schnitt erhielt die Beta-Version des Tests die Note 2. Die empfohlene Bearbeitungszeit (60 Minuten) stuften etwa gleich große Gruppen als „zu kurz“, „etwas zu kurz“, „genau richtig“ und „etwas zu lang“ ein - dies wird also hochgradig individuell empfunden. Feedbacks und Testergebnisse der Teilnehmer\*innen werden regelmäßig ausgewertet und dienen somit einem stetigen Verbesserungsprozess, der Ende 2020 in die Veröffentlichung der finalen Version des MINTFIT Informatiktests münden wird.

## **5 Zusammenfassung und Ausblick**

Der vorliegende Beitrag beschreibt den Informatiktest für MINT-Studieninteressierte des Verbundprojektes MINTFIT. Bei der Testerstellung wurde auf bestehende Lehrpläne, Hochschulcurricula, eine Dozent\*innenumfrage sowie die langjährige Erfahrung im Bereich E-Learning der Mitarbeiter\*innen zurückgegriffen. Seit Veröffentlichung der Beta-Version des Tests, die sowohl bei Fachwissenschaftler\*innen als auch der Zielgruppe überwiegend positives Feedback erhält, werden Nutzungsstatistiken und Feedbacks ausgewertet, um unklare Formulierungen, zu schwierige bzw. leichte Fragen zu identifizieren, anzupassen und somit den Test kontinuierlich zu verbessern. Zudem wird parallel an einem Informatik-Onlinekurs gearbeitet, der noch offene Fragen aus dem Testversuch klären und Interessierten Informationen zur vertiefenden Vorbereitung auf eine Ausbildung bzw. ein Studium im Informatik-Kontext bieten soll. Die Veröffentlichung des Kurses und der finalen Testversion wird im Jahr 2020 erfolgen. Anschließend ist geplant, mit vielfältigen Akteuren in Gespräche über einen bundesweit einheitlichen Mindestanforderungskatalog für Informatik im Spannungsfeld von Schulbildungsplänen und Hochschulcurricula einzutreten.




## 6 Danksagung

MINTFIT Hamburg ist ein Verbundprojekt der Hamburger MINT-Hochschulen Hochschule für Angewandte Wissenschaften Hamburg (HAW), HafenCity Universität Hamburg (HCU), Technische Universität Hamburg (TUHH), Universität Hamburg (UHH) sowie dem Universitätsklinikum Hamburg-Eppendorf (UKE) und wird gefördert von der Hamburger Behörde für Wissenschaft, Forschung und Gleichstellung.

## Literaturverzeichnis

- [Bi19] Bildungsstandards Informatik SI und SII, Gesellschaft für Informatik, <https://informatikstandards.de>, Stand: 18.06.2020.
- [BS18] Barbas, H.; Schramm, T.: The Hamburg Online Math Test MINTFIT for prospective Students of STEM Degree Programs. MSOR Connections, S. 43-51, 2018.
- [co14] cosh, cooperation schule:hochschule, <http://cosh-mathe.de/download/makV2.0neu.pdf>, Stand: 14.06.2020.
- [Do20] Dozent\*innen-Umfrage: Welche Informatik-Kompetenzen werden von Studienanfänger\*innen erwartet?, MINTFIT Hamburg, <https://www.mintfit.hamburg/ergebnisse-dozentinnenumfrage-informatik/>, Stand: 22.06.2020.
- [He10] Heublein, U. et.al.: Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen: Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08, HIS, Hannover, 2010.
- [He17] Heublein, U. et.al.: Zwischen Studiererwartungen und Studienwirklichkeit. Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote. DZHW, Hannover, 2017.
- [Me19] Meiling, S. et.al.: MINTFIT: Chemie-Onlinetest und -Kurs zur MINT-Studienvorbereitung am Übergang Schule-Hochschule. In (Meissner et.al.): Tagungsband zum 4. Symposium zur Hochschullehre in den MINT-Fächern, Nürnberg, S. 286-293, 2019.
- [MS19] Müller, U. C.; Sitzmann, D.: MINTFIT Physik: Onlineangebot zur Vorbereitung auf ein MINT-Studium. In (Nordmeier, V.; Grötzebauch, H. Hrsg.): Beiträge der DPG-Frühjahrstagung, Aachen, S. 287-290, 2019.
- [MW06] Maass, S.; Wiesner, H.: Programmieren, Mathe und ein bisschen Hardware ... Wen lockt dies Bild der Informatik? Informatik-Spektrum 29, S. 125-132, 2006.
- [Si19] Sitzmann, D. et.al.: Entwicklung eines Informatik-Onlinetests zur Studienvorbereitung im Projekt MINTFIT Hamburg. In (Meissner et.al. Hrsg.): Tagungsband zum 4. Symposium zur Hochschullehre in den MINT-Fächern, Nürnberg, S. 277-285, 2019.
- [Te20] Test Yourself!, Einstieg Informatik, [https://www.einstieg-informatik.de/index.php?article\\_id=800](https://www.einstieg-informatik.de/index.php?article_id=800), Stand: 18.6.2020.

## FL-Trail: Gruppenbasiertes forschendes Lernen digital unterstützen

Julian Dehne<sup>1</sup> , Sven Strickroth<sup>2</sup>  und Ulrike Lucke<sup>3</sup> 

**Abstract:** Forschendes Lernen stellt hohe Ansprüche an die Lehrenden und Lernenden. Basierend auf der Beobachtung, dass der Forschungsprozess trotz großer fachlicher Unterschiede sowohl in der Konzeptionsphase als auch in der Auswertungs- und Disseminationsphase gemeinsamen Ritualen folgt, wurde ein generischer Lehr-Lernprozess für das forschende Lernen modelliert. Dieser dient als Grundlage für ein in diesem Paper vorgestelltes integriertes, digitales Werkzeug zur Unterstützung des Ablaufs in Kursen mit forschendem Lernen-Ansatz. Dieser Ansatz verbindet Techniken aus dem Bereich des Computer-Supported Collaborative-Learning und dem E-Assessment und wurde mittels der dokumentarischen Methode qualitativ evaluiert. Die Ergebnisse zeigen eine Bandbreite von Gestaltungsempfehlungen sowie neuer Forschungswege und -formen auf.

**Keywords:** Forschendes Lernen, Gruppenorientierung, Bildungstechnologien

### 1 Motivation

Forschendes Lernen ist ein Ideal der Hochschuldidaktik. Wie projektorientiertes Lernen und problembasiertes Lernen lässt sich das forschende Lernen als Prinzip [Ba14] im Kanon didaktischer Theorien einordnen. Es durchdringt damit alle didaktischen Dimensionen wie z. B. das Verhältnis von Lernerorientierung oder auch die der medialen Methoden. Forschendes Lernen und der gezielte Einsatz von Medien in diesem Kontext stellen hohe Ansprüche an die Lehrenden und Lernenden. Für die Umsetzung von forschendem Lernen werden verschiedenste separate Tools eingesetzt [DLS17]. Eine integrierte (technische) Unterstützung für gruppen- oder kursorientierte Formate, die forschendes Lernen umsetzen, gibt es bisweilen nicht.

Trotz der inhärenten Komplexität ist es von doppelter Wichtigkeit forschendes Lernen und Bildungstechnologien in einen schlüssigen Bezug zu setzen. Zum einen hat forschendes Lernen den Status eines Goldstandards für die Hochschuldidaktik, da sowohl aus Bildungsperspektive die Schaffung einer akademischen Kultur hervorgehoben wird, aber auch mit der Brille der pädagogischen Psychologie fortgeschrittene Kompetenzen auf der Bloom'schen Taxonomie angesprochen werden. Zum anderen geben die Bildungstechnologien dem forschende Lernen die Chance, die Ideen zu erden,

---

<sup>1</sup> Universität Potsdam (UP), Institut für Informatik August-Bebel-Straße 89, 14482 Potsdam, julian.dehne@uni-potsdam.de, <https://orcid.org/0000-0001-9265-9619>

<sup>2</sup> (UP) sven.strickroth@uni-potsdam.de, <https://orcid.org/0000-0002-9647-300X>

<sup>3</sup> (UP) ulrike.lucke@uni-potsdam.de, <https://orcid.org/0000-0003-4049-8088>

da sie für eine technische Umsetzung auf ein Praxis-Niveau hin spezifiziert werden müssen, was für den Diskurs in der Bildungswissenschaft sonst nicht benötigt würde.

In diesem Paper wird ein integriertes Werkzeug zur digitalen Unterstützung von Kursen mit forschenden Lernen-Ansatz vorgestellt. Ziel des Ansatzes ist es, Lehrende und Lernende beim Prozess der Durchführung durch vorgegebene, anpassbare Strukturen zu unterstützen. Dies soll längerfristig zu verbesserten Kursstrukturen, aber auch zur Vereinfachung didaktischer Schlüsselprobleme führen. Dazu wurde zuerst ein allgemeines Prozessmodell von forschenden Lern-Abläufen erarbeitet, das schließlich durch das entwickelte System implementiert wird. Zur Evaluation kam eine qualitative Interviewstudie zum Einsatz, um einerseits die Eignung des vorgestellten Systems zu überprüfen und andererseits auch weitere Erkenntnisse darüber zu gewinnen, wie solche Unterstützungssysteme allgemein aussehen müssen.

## 2 Stand der Forschung

[HR19] definieren das forschende Lernen wie folgt:

*„Forschungsnahes Lernen umfasst über forschendes Lernen im engeren, in dem Studierende einen Forschungsprozess selbst forschend vollständig durchlaufen, hinaus alle diejenigen Formen des Lehrens und Lernens, welche die Studierenden explizit an Forschung als Prozess heranführen, indem sie einen solchen nachvollziehbar vor- und zur Diskussion stellen oder die Studierenden Elemente daraus als Ausschnitte aus einem mitgedachten Forschungszusammenhang üben und erlernen lassen.“* ([HR19], S. 3)

Mit dieser Definition wird das forschende Lernen auch auf Formen erweitert, die in der Nähe eigentlicher Forschung existieren. Die Grundidee, dass Studierende einen Forschungsprozess selbst forschend durchlaufen, auch wenn dieser dabei nicht vollständig praktisch umgesetzt wird, ist zentral.

[DLS17] haben den Versuch unternommen, den Stand der Nutzung von Medien zur Unterstützung von forschendem Lernen in Deutschland empirisch zu untersuchen. Sie kommen zu dem Ergebnis, dass es der Klärung methodologischer Fragen wie Definition und Reichweite des Medienbegriffs bedarf, um ein klares Bild des Nutzungsverhaltens zu zeichnen. Es gäbe vor allem eine unreflektierte Nutzung verbreiteter Werkzeuge, bei der das forschende Lernen, aber auch was ein Medium ausmacht, sehr unterschiedlich interpretiert würde. Nutzungsmustern von komplexeren Bildungstechnologien wurden vereinzelt im Bereich des Game-Based Learning und der virtuellen Realität gefunden.

[DKL19] haben die digitale Unterstützung von forschendem Lernen hinsichtlich der Möglichkeiten zur Reflexionsunterstützung untersucht. Reflexion, sowohl als Selbst-reflexion als auch als Feedback durch Peers, wurde als relevante Größe für das forschende Lernen identifiziert.

Weitere Ansätze zur Unterstützung der Kursorganisation sind in verwandten Bereichen wie dem Projekt-basierten Lernen und dem Computer-Supported-Collaborative-Learning (CSCL) zu finden. Es existieren Algorithmen zur Optimierung von Gruppenformation [Be17] und Projektablaufen [Mo16]. Technische Lösungen, die in der Lage sind lernerzentrierte Umgebungen gruppen- oder projektorientiert zu adaptieren, finden sich unter dem Begriff Personal-Learning-Environment (PLE) [KLZ14]. Es gibt Ansätze, um allgemein Kurse bzw. forschendes Lernen-Kurse bei spezifischen Aufgaben zu unterstützen. Jedoch müssen dazu zwingend verschiedene, unabhängige Tools genutzt werden und es gibt keine durchgehende Darstellung und Unterstützung des Prozesses. Diese Lücke wird von dem in diesem Paper vorgestellten System und vorgeschlagenen Prozessmodells geschlossen.

### 3 Konzeptentwicklung

Zur Prozessmodell-Entwicklung wurde die Arbeit von Dürnberger [Dü14] als Grundlage verwendet. Nach Dürnberger ist forschendes Lernen als eine dem Konstruktivismus zuzuordnende Lernform und umfasst: Bezüge zur Problemorientierung, produktives Lernen, Lernerzentrierung, projektorientiertes Lernen, Selbstorganisation, kritisch-reflexive Distanz und Soziale Kontextualisierung. Sie betont die Rolle, die Medien für die Konstruktion von Wissensinhalten spielen. Die didaktischen Dimensionen der Vermittlung, Aktivierung und Betreuung werden in Abbildung 1 dargestellt.

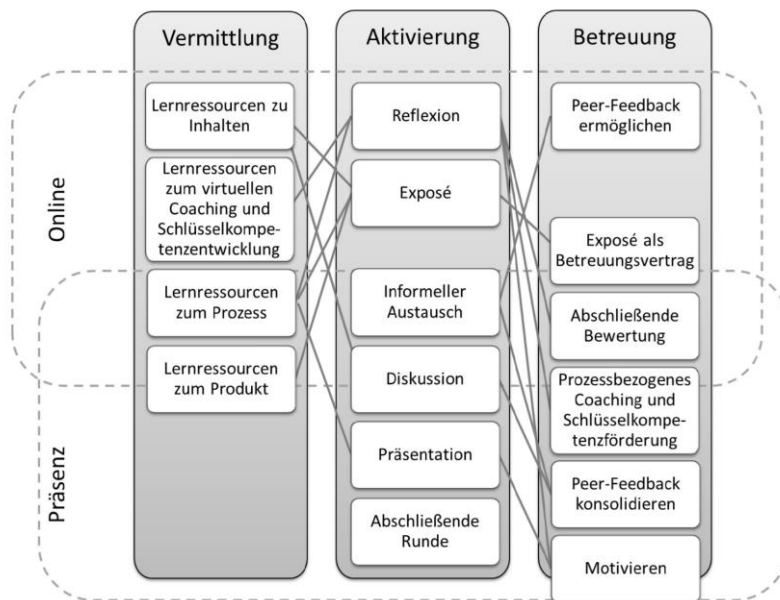


Abb. 1: Rolle der Medien für das forschende Lernen [Dü14]

Das Modell ist jedoch limitiert auf den Prozess des Schreibens einer Abschlussarbeit. Die Peers kommen als Feedback-Quelle vor, werden aber nicht in einem gemeinsamen Gruppenarbeitskontext gedacht, da die Abschlussarbeit üblicherweise allein geschrieben wird. Eine weitere Limitation ist die Modellierung des Forschungsprozesses, den die Studierenden durchlaufen müssen. Durch eine starke fachkulturelle Orientierung werden längere Entwicklungsphasen nicht berücksichtigt. Zuletzt unterschlägt der Fokus auf die Abschlussarbeit die Schwierigkeiten, offene Gruppenprozesse objektiv benoten zu müssen. Zusammenfassend lässt sich festhalten, dass in dem Modell von Dürnberger die Aufgabe der Medien wie auch der Prozess (Konzeptentwicklung, Peer Feedback, Produktion und weiteres Feedback) übernommen werden können. Erweitert werden muss der Ansatz um die Aspekte der Gruppenarbeit und der Bewertung.

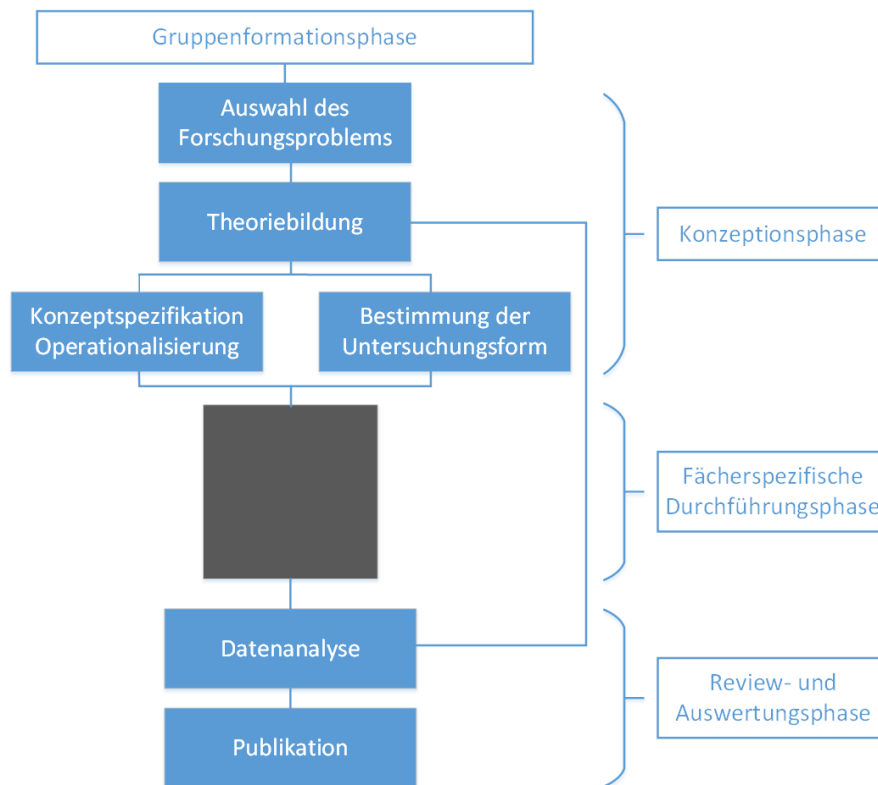


Abb. 2: Erweitertes Prozessmodell als Grundlage für die Entwicklung

In Abbildung 2 ist das entwickelte, erweiterte Prozessmodell dargestellt. Dieses ist generisch gehalten und enthält die vier Phasen *Gruppenformation*, *Konzeptionsphase*, *fächerspezifische Durchführungsphase* sowie eine *Review- und Auswertungsphase*, die sich typischerweise in forschenden Lernen Kursen finden.

In der Gruppenformationsphase werden die Arbeitsgruppen gebildet. Diese entwickeln in der Konzeptionsphase ein Forschungsdesign. Letzteres dient als Grundlage für eine erste Feedback-Runde. Daraufhin gehen die Gruppen in die Durchführung der Studie oder die Entwicklung. In dieser Phase kann vor allem die Reflexion unterstützt werden, wenn der fächerunabhängige Charakter gewahrt bleiben soll. Nach der Projektdurchführung folgt eine finale Feedback- und Bewertungsphase, in der diesmal nicht das Konzept, sondern das entwickelte Produkt bzw. die Studienergebnisse im Fokus stehen.

Die Kernidee besteht also in der Beobachtung, dass sich der Prozessanfang, das Finden der Gruppen und die Konzeptentwicklung *zwischen den Disziplinen* ähneln, wohingegen die Projektdurchführung kaum im Detail generisch unterstützt werden kann. Daher wurde dieser Schritt in Abbildung 2 als Blackbox dargestellt. Auch das Projektende läuft wieder in einer gemeinsamen Präsentations- bzw. Publikationskultur zusammen. Daher wurde Peer Review als Abschluss für die Konzeptphase und Peer Assessment für die finale Abgabe und Präsentation eingesetzt.

#### 4 Das System FL-Trail

FL-Trail wurde entwickelt, um den im vorherigen Abschnitt vorgestellten Prozessablauf in forschenden Lernen-Kursen sichtbar zu machen und zu unterstützen. Abbildung 3 zeigt einen Screenshot der Anwendung. Die Oberfläche ist aufgabenorientiert gestaltet. Die einzelnen Phasen und der Fortschritt werden graphisch auf der linken Seite sowie die erledigten und zu erledigenden Aufgaben auf der rechten Seite dargestellt. Bei den Aufgaben wird links oben ikonisch symbolisiert, ob die Aufgabe eine Gruppenaufgabe oder eine Einzelaufgabe ist. Rechts oben wird die Frist für die Aufgabe angezeigt. Die Aufgaben selbst bestehen aus einem beschreibenden Text sowie einer Schaltfläche mit einem Link. Für die vorgegebenen Phasen befinden sich dahinter standardmäßig integrierte Werkzeuge. Es können aber auch externe Werkzeuge eingebunden werden, die eine entsprechende Schnittstelle implementieren. Zusätzlich stehen die Produkte, hier Ergebnisse genannt, der einzelnen Phasen immer links unten zum Download bereit, sobald sie final eingereicht wurden. Allgemein bietet FL-Trail noch einen integrierten Chat mit automatisch angelegten Räumen für die Arbeitsgruppe sowie für den gesamten Kurs.

Der Prototyp führt die Lehrenden und Lernen durch die einzelnen Phasen, wobei der Lehrende die Phasenübergänge aktiv steuert. Das System informiert die Lernenden darüber automatisch per E-Mail und erinnert an offene Aufgaben. Die Phasen sind modular gehalten und können zum Teil übersprungen werden. Dies trifft z. B. auf die Entwurfsphase zu, in der die Studierenden ein Konzept für das Projekt ausarbeiten. Alternativ kann auch die Durchführungsphase (hier Projekt genannt) weggelassen werden und der Phasenübergang geschieht direkt zur Bewertung des Konzeptentwurfs. Somit sind Lehrende nicht strikt an das Prozessmodell gebunden.



Abb. 3: Ausschnitt der Benutzeroberfläche für Studierende

Für die Zuordnung der Peer Assessments und Peer Feedbacks wurde ein eigener Algorithmus entwickelt, der hier nicht vollständig erörtert werden kann. Dieser fasst gruppeninternes Feedback, aber auch Peer Feedback zwischen den Gruppen und der Bewertung der oder des Lehrenden zu einer Gesamtbewertung zusammen.

Zur Entkoppelung wurde eine Service-orientierte Architektur (SOA) mit REST-Schnittstellen gewählt. Die einzelnen Komponenten (FL-Trail, Rocket.Chat, GroupAI, Compbase [DL15], Lernreflex [DKL19]), aber auch deren benötigte Datenbanken (MySQL, Neo4j, ...) ließen sich als Docker-Images zusammenfassen. Die Rolle der Orchestrierung im SOA-Dreieck erfolgt mit Docker-Compose. Die entwickelte Software ist als Open Source verfügbar<sup>4</sup>.

## 5 Evaluation

Ziel der Evaluation war es, FL-Trail in einem qualitativen Experiment einzusetzen, um die Möglichkeiten der Unterstützung von forschendem Lernen zu explorieren und die bestehenden Konzepte weiterzuentwickeln. Zur Evaluation wurden semi-strukturierte Experteninterviews durchgeführt, die sowohl technische als auch didaktische Themen aufgriffen. Der Ablauf war so konzipiert, dass zuerst ein Cognitive Walkthrough [SB06] durchgeführt wurde, bei dem die Expertinnen und Experten die Rolle eines Lehrenden und eines Lernenden in einem Kurs einnahmen, in dem FL-Trail eingesetzt wird. Zur Durchführung der Studie wurde für die entwickelte Software eine Simulationsumgebung

<sup>4</sup> <https://gitup.uni-potsdam.de/fides/fltrail> – Zugriff am 22.06.2020



entwickelt, die es erlaubt, einen Kurs mit 30 Studierenden in seinem Ablauf durchzugehen. Für die Produkte und Kommunikation der Studierenden wurden Dummytexte generiert, so dass ein möglichst realistisches Bild eines Einsatzes des Werkzeugs möglich wurde. Das Verhalten bei den Eingaben, z. B. zögern bei unklarem Menü wurde mittels Bildschirmaufzeichnung mitgeschnitten. Durch das laute Denken konnten konzeptionelle Probleme des Werkzeugs von technischen Hürden unterschieden werden. Im Anschluss an den Durchlauf eines simulierten Semesters wurde das Interview geführt, das einzelne Probleme im Durchlauf aufgegriffen, aber auch fallübergreifende Fragen gestellt hat, um die Annahmenstruktur der Befragten zu beleuchten. Die Bildschirmaufzeichnung wurde mit den Audio-Aufnahmen zu einem Video zusammengeschnitten, so dass eine Grundlage für eine integrierte Betrachtung von Denken, Handeln und anschließender Reflexion in den Interviews gelegt wurde.

Als Expertinnen und Experten wurden aus dem Fachbereich Bildungstechnologien und der Deutschen Gesellschaft für Hochschuldidaktik sieben ProfessorInnen rekrutiert, die die Rolle eines Lehrenden in der Simulation einnehmen konnten und selbst schon Erfahrung mit forschenden Lernen-Kursen hatten. Als Auswertungsmethode wurde eine für Experteninterviews angepasste Variante der dokumentarischen Methode gewählt [No17]. Die Methode lässt sich in die Gruppe der qualitativen Methoden einordnen. Sie ist für das gewählte explorative Vorgehen gut geeignet, da sie nicht nur Textmaterial, sondern auch Ton und Bild betrachtet. Weiterhin erfüllt sie den Zweck, die Struktur der Annahmen während der Entwicklung (Vorstellungen von Softwaredesign, Kursdurchführung im forschenden Lernen, Interpretation von forschendem Lernen) mit dem dokumentierten Wissen der Expertinnen und Experten zu kontrastieren.

## 5.1 Ergebnisse

Im Folgenden wird ein Ausschnitt der Ergebnisse der explorativen Studie dargestellt. Es wurden während iterativer formulierender und reflektierender Interpretationen folgende fünf Schwerpunkte der Perspektiven in Cluster eingeteilt, bei denen sich die Annahmen zwischen einem pragmatischen und idealistischen Pol aufteilen.

### *Typus 1: Selbstreguliertes Lernen*

Der erste Typus basiert auf der Interpretation von forschendem Lernen auf Seiten der Lehrenden und deren Haltung zur Steuerung im didaktischen Prozess:

- Offen-idealistischer Typus: Forschendes Lernen bedeutet, die Vorstellung des Lehrenden als Regulativ aufzugeben.
- Pragmatisch-neutraler Typus: Forschendes Lernen ändert nichts an der Rolle der Lehrenden.
- Reguliert-traditioneller Typus: Forschendes Lernen schließt ein, dass der Lehrende gerade wegen der Komplexität in hohem Maße didaktisch aktiv werden muss.

*Typus 2: Definition forschenden Lernens*

Dieser Typus basiert auf der reflektierten Meinung der Expertinnen und Experten, die eine akademische Einstellung zu dem Begriff „forschendes Lernen“ haben:

- Typus breiter Definition: Forschendes Lernen lässt sich nicht als Archetyp für Lehr-Lern-Prozesse auffassen.
- Typus neutrale Definition: Es gibt verschiedene Archetypen für forschendes Lernen als Lehr-Lern-Prozess.
- Typus enge Definition: Forschendes Lernen bedeutet das Durchleben eines vollständigen Forschungsprozesses und strukturiert damit Lehr-Lern-Prozesse vor. Es lässt sich ein archetypischer Lehr-Lern-Prozess modellieren.

*Typus 3: Technik und Automatisierung*

In diesem Typus wird die Einstellung zu Technologien mit Blick auf didaktische Methoden zusammengefasst:

- Technik-kritischer Typus: Jede technische Lösung enthält in sich problematische Annahmen, die sich in dem Artefakt manifestieren.
- Technik-neutraler Typus: Technische Lösungen sind Werkzeuge, die abhängig von der didaktischen Methode eine sinnvolle Funktion haben können.
- Technik-affiner Typus: Es gibt Situationen im Lehr-Lern-Prozess, die sich mittels digitaler Medien effizienter und in manchen Fällen qualitativ neuartig angehen lassen.

*Typus 4: Didaktik als Spielfeld*

Dieser Typus ergibt sich aus der Einstellung zu didaktischen Innovationen in der Praxis.

Nur weil Forschende zur Hochschuldidaktik forschen, müssen sie nicht unbedingt dazu bereit sein, in ihrer eigenen Lehre zu experimentieren. Umgekehrt kann die wissenschaftliche Erkenntnis vor dem Wohl (einer betroffenen Kohorte) von Studierenden rangieren.

- Perfektionistisch-kritischer Typus: Didaktische Innovationen müssen eine hohe Hürde nehmen, bevor sie als praxisrelevant erachtet werden können.
- Innovativ-pragmatischer Typus: Eine didaktische Innovation kann in gesichertem Rahmen eingesetzt werden.
- Innovativ-experimenteller Typus: Eine didaktische Innovation kann direkt eingesetzt werden.

*Typus 5: Gruppenorientierung im forschenden Lernen*

Dieser Typus wurde aus den Diskussionen zum Thema Prüfen und Gruppenformation abgeleitet:

- Traditioneller Typus: Gruppenarbeit ist eine temporäre Form des Lernens, aber lernen und Prüfungen schreiben ist am Ende des Tages eine individuelle Angelegenheit.
- Pragmatischer Typus: Gruppen werden in Kursen aus ökonomischen Gründen gebildet, um die Bewertung besser bewerkstelligen zu können.
- Kollektiver Typus: Gruppenlernen ist die ideale Arbeitsform für das forschende Lernen, weil sie das Regulativ vom Lehrenden auf die Ebene der Peers überträgt.

Als einzelne Aussagen sollte aus den Interviews hervorgehoben werden, dass durch die klare Modellierung des Prozesses auf der einen Seite, aber auch durch die Automatisieren, bei mehreren Interviews das Gefühl eines Kontrollverlustes erzeugt wurde.

Bei den technischen Gestaltungsempfehlungen gab es noch weitere Details, die an dieser Stelle aus Platzgründen ausgelassen werden müssen. Auch die durchgeführte Studie zur Usability wird hier nicht weiter angeführt, da die qualitativen Punkte eine größere Rolle spielen und die Aussagekraft des System Usability Score zwar sehr positiv, aber bei einer einstelligen Zahl an Fällen nicht belastbar sind. In den Interviews benannte Usability-Probleme beziehen sich in den meisten Fällen auch auf die Komplexität bzw. Intransparent der Algorithmen und die Flexibilität des modellierten Lehr-Lernprozesses.

Im Folgenden werden in den Typen zusammengefassten Ansichten der Expertinnen und Experten sowie die eigenen Annahmen aus der Entwicklung abgetragen. Der minimale Kontrast wird mit dem Wert 1 belegt, der maximale mit dem Wert 3. Daraufhin wird aus den Differenzialen der Expertinnen und Experten der Durchschnitt gebildet, der daraufhin gegen die Annahmenstruktur des Autorenteams aus einer Selbstbefragung abgetragen wird.

Abbildung 4 zeigt, dass entgegen der Intuition die Expertinnen und Experten teilweise idealistischere Annahmen an die Technik oder das forschende Lernen haben. Das qualitative Vorgehen hat eine Reihe reichhaltiger Erkenntnisse und Empfehlungen für das weitere Vorgehen geliefert, die im Folgenden vorgestellt werden:

1. Adressierung der Bedenken gegenüber der Automatisierung
2. Stärkere Visualisierung komplexer Prozesse, insbesondere bei der Gruppenbildung
3. Einbau von Rückschritten in den Zustandsautomaten des Prozessmodells
4. Adressierung des Problems von Abbrechern
5. Ausbau der Durchführungskomponente in Bezug auf konkrete Forschungsformen
6. Einsatz der Reflexionsunterstützung über den gesamten Kursverlauf

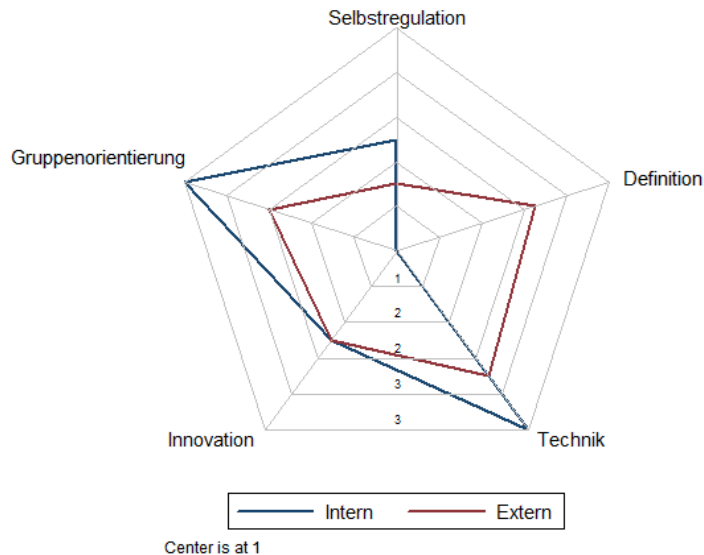


Abb. 4: Ergebnis der sinngenetischen Typenbildung

Basierend auf diesen Gestaltungsempfehlungen ergeben sich folgende Vorschläge für eine technische Fortführung:

1. Weiterentwicklung der aufgabenorientierten Architektur hinsichtlich Flexibilisierungsanforderungen z. B. mehr Freiheiten bei der Prozessgestaltung
2. Entwicklung von LMS-Plugins als Alternative zur prototypischen Oberfläche
3. Anbindung einer mobilen Applikation und/oder Gestaltung einer mobilen Oberfläche

Diese Hinweise wurden in den Interviews mehrfach benannt und sollten als Ergebnis daher auch überliefert werden. Insgesamt wurde das Werkzeug in den meisten Fällen als zum Einsatz geeignet eingestuft und kann nicht durch bestehende Systeme ersetzt werden (diese Fragen wurde in den Interviews explizit gestellt).

Das generische Prozessmodell wurde in den Interviews bestätigt. Es konnten trotz Übersicht der Expertinnen und Experten über verschiedene Disziplinen keine Beispiele gefunden werden, in denen das Modell angewendet werden konnte. Zudem wurde die Argumentation für das Prozessmodell in der Argumentationsanalyse mit positivem Ergebnis überprüft.

## 6 Diskussion und Schlussbetrachtung

In der Evaluation zeigt sich der Wert der qualitativen Herangehensweise. Hätte eine evidenzbasierte Bewertung stattgefunden, wären die Annahmen unreflektiert eingegangen, dass die Experten der Automatisierung vormals manueller Prozesse wie dem Peer-Assessment oder der automatisierten Gruppenformation zustimmen, oder die dadurch erzielten Effizienzgewinne schätzen. Tatsächlich erzeugen die verwendeten Algorithmen einen gefühlten Kontrollverlust, da als wichtig eingestufte Entscheidungen als intransparent wahrgenommen werden. Das Design-based Research (DBR)-Vorgehen sieht daraufhin eine weitere Iteration vor, in der sowohl die Software als auch der didaktische Prozess so angepasst werden, dass Vertrauen hergestellt werden kann. Die Frage bleibt jedoch offen, ob es eine Grenze gibt, an der die (empfundene) Komplexität der Technik und die Affinität zu Technik im Prinzip nicht durchbrochen werden können, so dass eine Weiterentwicklung für den allgemeinen Nutzer keinen mehr Sinn ergibt.

DBR und die dokumentarische Methode stellen eine wirkungsvolle methodologische Verbindung dar, um Bildungstechnologien im Sinne eines qualitativen Experiments einzusetzen. Abhängig von der Komplexität der Entwicklung und dem Stand der Entwicklung eignet sich die explorative Herangehensweisen, um in der Fülle der zu betrachtenden medien-didaktischen Faktoren, institutionellen Beschränkungen und psychologischen Fakten die entscheidenden Punkte zu finden, die für den Erfolg oder Misserfolg einer Bildungsinnovation relevant sind. Dies schließt evidenzbasierte Methoden nicht aus, sondern bildet vielmehr die Grundlage für den sinnhaften Einsatz dieser. In dem Fall des hier entwickelten Werkzeuges sind die nächsten Schritte, die Desiderata, sowohl technischer als auch konzeptioneller Natur, zu erforschen und weiterzuentwickeln. Bei einer größeren Verbreitung könnten dann auch empirische Studien folgen, um den komparativen Nutzen gegenüber bisheriger Lehr-Lernpraktiken und alternativen digitalen Werkzeugen festzustellen.

Es konnte gezeigt werden, dass das forschende Lernen mit Bildungstechnologien direkt unterstützt werden kann. Es muss aber auch darauf hingewiesen werden, dass es von dem Begriff des forschenden Lernens abhängt. Bei einem engeren Begriff, der das Durchlaufen des vollständigen Forschungsprozesses vorsieht, passt das entwickelte Werkzeug besser als bei weiter gefassten Interpretationen, die auch forschungsnahes oder forschungsorientiertes Lehren und Lernen mit einschließen.

## Literaturverzeichnis

- [Ba14] Baumgartner, P.: Taxonomie von Unterrichtsmethoden: ein Plädoyer für didaktische Vielfalt, 2014.
- [Be17] Bellhäuser, H.; Konert, J.; Röpke, R.; Rensing, C.: Eine extravertierte und eine gewissenhafte Person in jeder Lerngruppe! Effekte der Verteilung von Persönlichkeitsmerkmalen auf Zufriedenheit und Lernergebnis. In: Proc. DeLFI 2017. S. 309–320, 2017.
- [DKL19] Dehne, J.; Knoth, A.; Lucke, U.: Studieneingangsphase stärken, forschend lernen und digitale Medien nutzen. In: Forschendes Lernen in der Studieneingangsphase. S. 111–125, 2019.
- [DL15] Dehne, J.; Lucke, U.: An infrastructure for cross-platform competence-based assessment. In: CHANGEE Facing the Challenges of Assessing 21st Century Skills in the Newly Emerging Educational Ecosystems. S. 29–38, 2015.
- [DLS17] Dehne, J.; Lucke, U.; Schiefner-Rohs, M.: Digitale Medien und forschungs-orientiertes Lehren und Lernen – empirische Einblicke in Projekte und Lehrkonzepte. In: Bildungsräume. S. 71–83, 2017.
- [Dü14] Dürnberger, H.: Forschendes Lernen unter Einsatz digitaler Medien beim Verfassen der Bachelorarbeit: Potenziale für die Schlüsselkompetenzentwicklung, 2014.
- [HR19] Huber, L.; Reinmann, G.: Vom forschungsnahen zum forschenden Lernen an Hochschulen, 2019.
- [KLZ14] Kiy, A.; Lucke, U.; Zoerner, D.: An adaptive personal learning environment architecture. In: International Conference on Architecture of Computing Systems. S. 60–71, 2014.
- [Kö13] Konert, J.; Burlak, D.; Göbel, S.; Steinmetz, R.: GroupAL: ein Algorithmus zur Formation und Qualitätsbewertung von Lerngruppen in E-Learning-Szenarien mittels n-dimensionaler Gütekriterien. In: Proc. DeLFI 2013. S. 71–82, 2013.
- [Mo16] Monett, D.; Kiehne, B.: Interdisziplinäres Projektlernen in der agilen Softwareentwicklung. die hochschullehre 2(2), S. 1–20, 2016.
- [No17] Nohl, A.: Interview und Dokumentarische Methode. Anleitungen für die Forschungspraxis. 2017.
- [Re19] Reinmann, G.: Entwicklung als Forschung? Gedanken zur Verortung und Präzisierung einer entwicklungsorientierten Bildungsforschung. In (Reinmann, G., Hrsg.): Reader zu DBR. S. 70–82, 2019.
- [SB06] Sarodnick, F.; Brau, H.: Methoden der Usability Evaluation. 2006.

## Eine forschungspraktische Perspektive auf xAPI-Registries

Matthias Ehlenz<sup>1</sup>, Birte Heinemann<sup>1</sup>, Thiemo Leonhardt<sup>2</sup>, René Röpke<sup>1</sup>, Vlatko Lukarov<sup>1</sup> und Ulrik Schroeder<sup>1</sup>

**Abstract:** In der Learning Analytics-Forschung hat sich xAPI als Standard etabliert. Actor – Verb – Object. Während Idee einfach und elegant ist, ist die Umsetzung deutlich komplexer. xAPI stellt eine Grammatik, doch das Vokabular ist nicht eindeutig definiert. Extensions, Profiles, Recipes, Attachments, Objekte und Activities, viele Konzepte sind spezifiziert, die inhaltliche Ausgestaltung aber bewusst offengehalten. Die resultierende Konsequenz ist fehlende Eindeutigkeit. Dieser Beitrag arbeitet diese Problematik strukturiert auf und diskutiert die gegenwärtigen Ansätze verschiedener Registries durch die kritische Betrachtung der Entwicklung des xAPI Standards.

**Keywords:** xAPI Registry, Learning Analytics, Kollaboration, Transparenz, Open Science.

### 1 Motivation

Etablierte Modelle und Spezifikationen wie z. B. SCORM, spezifische systemgenerierte Daten (z. B. Logs) oder Forschungsmodelle, z. B. Learning Context Data Model [Th14] werden genutzt, um relevante Informationen zu teilen. Jedes dieser Werkzeuge stellt unimodal eine sinnvolle Option dar, bedient aber nicht die Anforderungen, die moderne Forschungsszenarien mit sich bringen. Diese erfordern multi-modale Datenquellen. Im Kontext der zunehmenden Bedeutung von Open Science soll Forschung nachvollziehbar und reproduzierbar sein und ein zentraler Teilbereich ist Open Data [Kr11].

Im Bereich Learning Analytics und Educational Data Mining sind zwei Datenformate etabliert: xAPI und IMS Caliper. Die Auswahl ist nicht Gegenstand dieses Beitrags. Die beteiligten Institutionen setzen auf xAPI, die ausschlaggebenden Argumente für xAPI lassen sich mit der Offenheit des Standards und den Schwerpunkten Beyond Browser, Distributed Learning & Multi-Agent Statements zusammenfassen [IM16]. Gegenstand dieses Beitrags ist demnach der praktische Einsatz des xAPI Datenformats speziell mit dem Fokus auf wissenschaftliche Kollaboration.

---

<sup>1</sup> RWTH Aachen, Informatik 9 (Learning Technologies), Ahornstr. 55, 52074 Aachen, {ehlenz, heinemann, roepke, lukarov, schroeder}@informatik.rwth-aachen.de

<sup>2</sup> TU Dresden, Professur für Didaktik der Informatik, Nöthnitzer Str. 46, 01187 Dresden, thiemo.leonhardt@tu-dresden.de

## 2 Grundlagen xAPI und Prozess der Standardisierung

Die Entwicklung der Spezifikation xAPI ist durch die Advanced Distributed Learning Initiative (ADL), einer Initiative des US-amerikanischen Department of Defence (DoD) angestoßen worden. Seit dem ersten Entwurf hat die Spezifikationen einige Veränderungen durchlaufen. Aktuell gibt es mit Version 1.0.3. keine Bestrebungen einer Weiterentwicklung [Jo16]. Das grundlegende Prinzip von xAPI hat sich im Laufe der Entwicklung nicht verändert: Ein Statement besteht aus dem Tripel {Actor, Verb, Object} bzw. *Wer macht Was Womit*. Im Laufe der Entwicklung wurden jedoch die genauen Modalitäten der Spezifikation der Elemente angepasst. Dies betrifft insbesondere die Definition der Verben. Ein Beispiel hierfür ist die Anzahl der reservierten Verben, aber auch die Definition der Verben durch eindeutige Identifier.<sup>3</sup>

Des Weiteren gab es bereits in einer frühen Phase der Entwicklung von xAPI die Idee, Definitionen für bestimmte Aufgaben und Bereiche zu vereinheitlichen und zusammenzufassen. Hierzu das Konzept der Recipes<sup>4</sup> (dt. Rezepte) eingeführt: Sammlungen von Verben, Aktivitäten, Erweiterungen, Regeln und Schemata, die beispielsweise die Erfassung der Interaktion eines Lernenden mit einem Video beschreiben. Der Begriff der Recipes findet sich bereits in der Version 1.0.0 in 2013 nicht mehr, da er sich als unpräzise herausstellte [Mi18]. Abgelöst wurden sie von Profilen formalisiert im JSON-LD-Format. Profile erfüllen den gleichen Zweck wie Recipes, bringen neben "Menschenlesbarkeit" auch Maschinenlesbarkeit mit.

### 2.1 Methodisches Vorgehen

Der Umfang existierender Publikationen und zuverlässiger Online-Quellen, insbesondere zur praxisnahen Implementation von Registries und Erfahrungsberichten der praktischen Nutzung, mit wenigen Ausnahmen, z.B. [Bl18, Ba16], ist begrenzt. Der Ansatz über wissenschaftliche Publikationen wurde deshalb auf Gebiete außerhalb wissenschaftlicher Publikationen nach folgendem Vorgehen über verschiedenen Suchmaschinen (Google, Bing, Yandex) erweitert: (1) Suche (xAPI|"Experience API"|"TinCan API") + (Registry|Repository|Verbs), (2) Suche nach entsprechenden Publikationen, (3) Sichtung aller offiziellen Dokumente zu xAPI und TinCan und (4) die Betrachtung exemplarischer Statements in Publikationen<sup>5</sup>.

Es existieren viele Verweise auf private Server, Blogeinträge, Firmen-Websites, Webinare und Git-Repositoryn. Die Analysekriterien umfassen (a) Aktivität und Pflege(zeitraum), (b) Anzahl der aktiv beitragenden Institutionen, (c) Transparenz der Prozesse für Beiträge von Forschern und Praktikern, sowie (d) der Umfang, also die

<sup>3</sup> Gleiches gilt für Activities und Extensions und ist im Folgenden im Sinne der Lesbarkeit auf das Kernkonstrukt Verb reduziert.

<sup>4</sup> <https://xapi.com/recipes>, zuletzt abgerufen am 16.4.2020

<sup>5</sup> Hierzu wurden z.B. IRIs (auch aus Screenshots) betrachtet und die URL-Parts im Browser getestet.



Anzahl der Verben und Aktivitäten, die innerhalb der Registry existieren. Darüber hinaus wurde der (e) Zustand der Registry betrachtet, d.h. die Anzahl der Definitionen mit vollständiger Dokumentation, die Anzahl und Umfang der dokumentierten Verben und Aktivitäten sowie der inhaltliche Kontext der Verben (militärische vs. zivile Nutzung).

## 2.2 Betrachtung existierender Ansätze

Die Rechercheergebnisse lassen sich in vier Kategorien einteilen: Erstens die "offiziellen" Repositories [registry.tincanapi.com/](http://registry.tincanapi.com/) und <http://xapi.vocab.pub/>. Die zweite Kategorie bilden die Repositories der Industrievertreter im Konsortium, von denen hier exemplarisch Rustici-Software (keine Updates in den vergangenen acht Jahren) und Brindleway (die exakt sieben Verben listen) genannt werden sollen, welche wegen der mangelnden Aktualität nicht tiefer eingegangen wird. Die dritte Kategorie stellen die verteilten Profile der CoPs dar. Die vierte Kategorie sind Ansätze aus der Wissenschaft heraus. Diese entwickeln sich meist lokal. Das Ende der Weiterentwicklung lässt sich dabei nicht immer reproduzieren. Bindung an Projektlaufzeiten, das Ausscheiden treibender Projektpartner, nicht-nachhaltige Technologiewahl oder schwierige Organisationsstrukturen, könnten Gründe für das Ende einer Entwicklung sein.

Die Registry [registry.tincanapi.com](http://registry.tincanapi.com/)<sup>6</sup> beinhaltet 182 Verben von insgesamt 10 Beitragenden<sup>7</sup>. Bei 125 dieser Verben sind die IRIs keine URLs, 117 hiervon resultieren in einem HTTP-Statuscode 404<sup>8</sup>, 4 zeigen auf unregistrierte Domains, die übrigen 4 sind Redirects auf Produkthomepages der beitragenden Softwarefirmen. Von den 57 URL-IRIs lösen 50 in Spezifikations-gemäßem JSON auf, der neben den Attributen "name" und "description" keine weiteren Metainformationen enthält. Die übrigen 7 bieten identischen Informationsumfang, aufbereitet in einer HTML-Tabelle. Von 113 Activities lösen 64 von 2 der 6 Beitragenden auf ohne Metainformationen. Gemäß Spezifikation dient die IRI der Distinktion des Kontextes. Aber lediglich zwei Verben nutzen die IRI zu diesem Zweck: entered frame und exited frame, die aber auch insofern vom Konzept abweichen, dass sie den Teilbegriff frame nicht nur in IRI und description kodieren, sondern es auch im display property (und damit im Namen) kodieren. Somit ist die IRI im Grunde nicht mehr zur Unterscheidung erforderlich.

Eine umfangreichere Datenbank ist [http://xapi.vocab.pub](http://xapi.vocab.pub/)<sup>9</sup>. Diese beinhaltet 538 Einträge mit teils wenig objektiv messbaren Verben wie "imagined" und "troubleshoot", als auch ein breites Spektrum militärischer Begrifflichkeiten wie "assaulted", "attacked", "triaged", "camouflaged" und "war-gamed". 465 IRIs lösen auf eine Informationsseite auf, die Namen und IRI enthält, jedoch nur 46 enthalten die Description aus der Übersicht. Die vorhandenen Descriptions bestehen im Mittelwert aus 225,5 Zeichen.

<sup>6</sup> laut ahref.com: 562 Backlinks von 52 Domains.

<sup>7</sup> gemäß des organizational Teils der IRIs.

<sup>8</sup> 88 hiervon sind Verweise auf activitystrea.ms, gemäß Spezifikation keine URLs

<sup>9</sup> zum Vergleich: 7333 Backlinks von 20 Domains

Dies weist auf einen tatsächlichen Erklärungsversuch der Autoren für die eingeführten Begriffe hin. Zusätzlich werden Hinweise zu passenden Extensions angegeben, so dass ein detaillierteres Bild möglich wird. Die IRIs enthalten semantische Hinweise auf den Kontext, die aber Hintergrundwissen verlangen, um wirklich interpretierbar zu werden: adb deutet auf “actionable data book”, ein Profil zur Erfassung der Interaktion mit Ebooks hin, dod-isd lässt auf Profile im Auftrag des DoD schließen. Die Verben des dod-isd-Profiles machen mit 423 einen Großteil des Bestands aus, geht man von der Spezifikations-gemäßen semantischen Eindeutigkeit der Definitionen aus verbleiben in dieser Registry 115 Verben zur zivilen Nutzung.

Ein von einer Interessensgemeinschaft chinesischer xAPI-Anwender geprägter Ansatz<sup>10</sup> ist seit knapp zwei Jahren ohne Aktualisierung, zeigt die Liste jedoch mehr Metainformation als alle bisher betrachteten Registries. Ein weiterer Ansatz kommt aus den Niederlanden: Zum Ansatz der TrustedLA gibt es vielversprechende Publikationen, das zugeordnete Repository<sup>11</sup> weist abgesehen von einem exemplarischen Statement keinen Inhalt und keine Aktualisierung seit 2018 auf. Vielversprechend dagegen sind die zugehörigen Dokumente auf Google Docs<sup>12</sup>. 2015 angelegt, stellen sie eine weit fortgeschrittene Ausarbeitung von xAPI-Semantik für die gelisteten Projekte dar. Über eine Fortführung der Arbeit ist nichts bekannt und seit 2018 sind keine weiteren Publikationen in diesem Zusammenhang erschienen. [Be16a, Be16b]

### 2.3 Schlussfolgerungen für die Praxis

Die **verschiedenen Stakeholder** erschweren den Umgang mit dem Standard. Insbesondere die Beteiligung von Dienstleistern mit Profitinteresse [BS18], und die Sonderrolle der ADL als Organisation, die dem DoD untersteht, beschränken bis auf wenige Ausnahmen, siehe [Be16b], die Entwicklung von Konventionen und Prozessen in Forschungsgemeinschaft. Infolge der für die Wissenschaft **ungewohnten Publikationsstrategien** entstehen Missverständnisse, die sich im Wording (Experience API, xAPI, Tin Can API, TCAPI) und der Auslegung des Standards widerspiegeln.

Die Spezifikation definiert das Format in dem Daten ausgetauscht werden sollen. Konventionen werden empfohlen, aber nicht vorgegeben. Die ursprüngliche Spezifikation sah mehrere reservierte Verben vor ([Bo12]). Seit Version 1.0.3. wird nur noch ein Verb (“void”) reserviert, um den angestrebten Lifecycle mit Vollständigkeit und Unveränderlichkeit praktisch umsetzen zu können. Alle weiteren Verben sind frei definierbar, um API im Sinne von “Beyond Browser” anzuwenden.

Die einzige Bedingung, die der Standard stellt, ist die Eindeutigkeit der Identifier. In Version 0.95 zeigte ein Identifier auf eine Ressource, die Meta-Informationen enthält:

<sup>10</sup> <http://xapi-cop.net/vocab/en>, abgerufen am 17.4.2020

<sup>11</sup> <https://github.com/TrustedLA/xAPI-Dutch-Spec>, abgerufen am 17.4.2020

<sup>12</sup> <http://bit.ly/DutchXAPIreg> und <http://bit.ly/DutchXAPISpread>, beide abgerufen am 17.4.2020

“For each new verb you must define what it means. Each time it’s used in a statement you have to point to that definition (with a URI – Uniform Resource Identifier)” [Bo12]. Diese Beschreibung hat sich geändert und aus URIs sind IRIs (Internationalized Resource Identifier) geworden. Im Fokus der aktuellen Spezifikation liegt nun die Bestimmung des Kontextes des Verbs durch die IRI. Nach wie vor ist es aber möglich (und Teil der Spezifikation) “hosted Metadata” zu hinterlegen, diese Möglichkeit wird in der offiziellen Registry<sup>13</sup> aber nur für 31,3% (57 von 182) der Verbdefinitionen genutzt.

Die Spezifikation beabsichtigt **Eindeutigkeit**, d.h. dass die Verb-ID <http://activitystrea.ms/schema/1.0/accept> eindeutig auf eine “akzeptierende” Interaktion zurückzuführen ist, aber nicht jede “akzeptierende” Interaktion auf diese ID hinausläuft. Während dies für unmittelbare Stakeholder wie Firmen und das US-amerikanische Militär in zahlreichen Anwendungsfällen ausreichend ist, ist es aus Perspektive der Wissenschaft problematisch. Die fehlende Eineindeutigkeit, d.h. umkehrbare Eindeutigkeit führt zu Problemen bei der Aggregation von Daten verschiedener Kontexte. Diese stellt mit der Vergleichbarkeit ein Schlüsselement des Open Data-Ansatzes dar.

Ein weiteres Problem in interdisziplinären, multimodalen Kontexten ist, dass deren **vielfältige Anforderungen** aktuell kaum berücksichtigt werden. Es gibt Profile für xAPI in VR, vermutlich auch eines, welches Eye-Tracking behandelt, für die Kombination beider Ressourcen konnten wir noch kein Profil finden. Und auch wenn Profile im Sinne von Open Data einen Fortschritt darstellen können, sind sie doch kritisch zu betrachten. Zum einen hält sich die Menschenlesbarkeit von JSON-LD in Grenzen. Die richtigen Vokabeln zu finden ist ohne entsprechende Aufbereitung problematisch. Definitionen der Verben sind “nested”, eingebettet im Profil, die Identifier folgen häufig dem Prinzip “IRIs are not URLs” und finden sich nur teilweise in einer der Registries wieder.

Zum anderen überlässt es die ADL sogenannten Communities of Practice (CoP), eigene Profile für ihre Domänen zu erarbeiten, was weiter zur **Fragmentierung der Lösungsansätze** beiträgt. Viele dieser CoPs haben Profile erarbeitet, sich vernetzt<sup>14</sup> und ihre Ergebnisse veröffentlicht. Eine Anlaufstelle auf der Suche nach dem passenden Profil ist das von der ADL gepflegte Repository<sup>15</sup> der Authored Profiles. Wer, wie die Autoren dieses Beitrag z.B. auf der Suche nach Profilen für Kollaboration an Multi-Touch Tabletop-Displays ist, wird hier allerdings nicht fündig. Eine Mitwirkung läuft via ADL und ist weder transparent noch offen. Das “offene” Profile-Repository<sup>16</sup> von ADL liefert lediglich die Spezifikation von Profilen und Details bis hin zur Vorgabe der server-seitigen Abfragesprache.

---

<sup>13</sup> <https://registry.tincanapi.com/>, abgerufen am 16.04.2020

<sup>14</sup> <https://xapi-cop.net/>, abgerufen am 17.4.2020

<sup>15</sup> <https://github.com/adlnet/xapi-authored-profiles>, abgerufen am 16.4.2020

<sup>16</sup> <https://github.com/adlnet/xapi-profiles/>, abgerufen am 16.4.2020

### 3 Schluss und Ausblick

In diesem Beitrag wurde der gegenwärtige Stand von xAPI systematisch aufbereitet. Durch die kritische Betrachtung der Entwicklung des Standards wurden Probleme identifiziert und diskutiert. Mit xAPI hat sich zwar ein praxistauglicher Standard entwickelt, jedoch führen fehlende Eindeutigkeit und eine eingeschränkte Offenheit insbesondere in interdisziplinären Projekten zu Problemen. Um langfristig einen offenen, unabhängigen und zielgerichteten Einsatz von xAPI sicherzustellen, braucht es den Konsens der Forschungscommunity und einen gemeinschaftlichen Ansatz zur Definition von Konventionen statt projekt- oder institutionsspezifischer Insellösungen.

#### Literaturverzeichnis

- [Ba16] Bakharia, A. et al.: Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. In (Gašević, D.; Lynch, G. Hrsg.): Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. Association for Computing Machinery, New York; S. 378-382, 2016.
- [Be16a] Berg, A. et al.: Dutch Cooking with xAPI Recipes: The Good, the Bad, and the Consistent. In (Chen, N.-S. et al. Hrsg.): 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT). IEEE; S. 234-236, 2016.
- [Be16b] Berg, A. et al.: The dutch xAPI experience. In (Gašević, D.; Lynch, G. Hrsg.): Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. Association for Computing Machinery, New York; S. 544-545, 2016.
- [Bl18] Blake-Plock, S.: xAPI: A Guide for Technical Implementers, New York, 2018.
- [Bo12] Bowe, M.: What's Up with Verbs in 95? <https://xapi.com/blog/whats-up-with-verbs-in-95>, Stand: 17.04.2020.
- [BS18] Bowe, M.; Silvers, A. E.: US DoD xAPI Profile Server Recommendations. Report, Narberth, 2018.
- [IM16] IMS Global Learning Consortium: Initial xAPI/Caliper Comparison. <http://www.imsglobal.org/initial-xapicaliper-comparison>, Stand: 17.04.2020.
- [Jo16] Johnson, A.: Interview with Andy Johnson about Newest Version of xAPI, version 1.0.3, 2016. <https://adlnet.gov/news/2016/09/21/interviewing-xapi-spec-group-organizer>, Stand: 17.04.2020.
- [Kr11] Kraker, P. et al.: The case for an open science in technology enhanced learning. International Journal for Technology Enhanced Learning, 3, 2011.
- [Mi18] Miller, B.: Profile Recipes vs. xAPI Profiles. <https://xapi.com/blog/profile-recipes-vs-xapi-profiles>, Stand: 16.04.2020.
- [Th14] Thüs, H. et al.: Kontexterfassung, -modellierung und -auswertung in Lernumgebungen. In (Trahasch, S. et al. Hrsg.): DeLFI 2014 - Die 12. e-Learning Fachtagung Informatik. Gesellschaft für Informatik e.V, Bonn; S. S. 157–162, 2014.

## Nutzung von GitHub für Open Educational Resources

### Analyse zu Prozessen der Versionsverwaltung

Nadine Schröder<sup>1</sup>, Peter Pfänder<sup>2</sup>

**Abstract:** Versionsverwaltung ist in der Softwareentwicklung weit verbreitet, wobei durchaus Potenziale für den Einsatz in anderen Bereichen bestehen, beispielsweise bei der Erstellung, Bearbeitung und Bereitstellung von Open Educational Resources (OER). Dieses Forschungsvorhaben untersucht in welcher Art und Weise GitHub als Plattform zur Versionsverwaltung zur Ablage und Nutzung von Lehr-/Lernmaterialien genutzt wird. Dabei finden insbesondere die Prozesse und Möglichkeiten der Versionsverwaltung Berücksichtigung, um der Frage nachzugehen, wie diese zu didaktischen Zwecken im Umgang mit Bildungsmaterialien genutzt und bei der kollaborativen Erstellung und Bearbeitung vorgegangen wird.

**Keywords:** Open Educational Resources, Infrastruktur, Versionsverwaltung, GitHub, Bildungsmaterialien, Lehr-/Lernmaterialien, Kollaboration

## 1 Einleitung und Hintergrund

In der Softwareentwicklung gilt Versionsverwaltung seit vielen Jahren als Standard, um Prozesse des gemeinsamen Arbeitens zu erleichtern und zu optimieren. Git ist ein etabliertes System zur verteilten Versionsverwaltung, auf dem die weit verbreitete webbasierte Plattform GitHub zur Bereitstellung von Softwarecode basiert. Versionsverwaltung dient zur Erfassung von Änderungen an Dokumenten und Dateien. Nachvollziehbarkeit ist vor allem bei kollaborativen Arbeitsprozessen innerhalb der Softwareentwicklung wertvoll, kann aber auch für die Erstellung und Bearbeitung von Lehr-/Lernmaterialien nützlich sein. Somit bietet sich GitHub auch als Plattform für die Ablage und den Austausch von Bildungsmaterialien an [Za15].

In diesem Forschungsvorhaben wird untersucht, in welcher Art und Weise GitHub für kooperative Erstellung, Bearbeitung und Bereitstellung von Lehr-/Lernmaterialien bereits genutzt wird. Ein Schwerpunkt liegt dabei darauf, inwieweit die Funktionalitäten und Workflows der Versionsverwaltung Anwendung finden, um Anforderungen für den Einsatz mit Bildungsmaterialien zu ermitteln. Die Analyse ist integriert in ein Forschungsvorhaben, bei dem untersucht wird, wie mit Versionen umgegangen werden kann, die bei

---

<sup>1</sup> Universität Duisburg-Essen, Learning Lab, Universitätsstr. 2, 45141 Essen, nadine.schroeder2@uni-due.de, <https://orcid.org/0000-0001-6650-4656>

<sup>2</sup> Universität Duisburg-Essen, Learning Lab, Universitätsstr. 2, 45141 Essen, peter.pfaender@uni-due.de

einer dezentralen und kooperativen Nutzung und Bearbeitung von Open Educational Resources (OER) entstehen. Das übergeordnete Forschungsprojekt beschäftigt sich mit einer verteilten Infrastruktur, um die Vernetzung und Verteilung von Open Educational Resources unter den deutschen Hochschulen zu fördern. Durch die Vernetzung von verschiedenen Hochschulsystemen über ein zentrales System und den Austausch von Metadaten sollen Auffindbarkeit und Zugriff auf Ressourcen standortübergreifend ermöglicht werden [Ke19].

## 2 Versionsverwaltung und Open Educational Resources

Die Idee von Open Educational Resources (OER) bezieht sich auf den Zugang zu Lehr-/Lernmaterialien nach dem Konzept von Open Science, der Forderung nach der Öffnung von Bildung und Wissenschaft, unter dem auch Open Source gefasst wird [FF14]. OER beruhen auf dem Konzept der 5 R, die verdeutlichen, welche Möglichkeiten offene Lizenzen beim Umgang mit Materialien bieten [Wi14]:

*Retain – the right to make, own, and control copies of the content*

*Reuse – the right to use the content in a wide range of ways*

*Revise – the right to adapt, adjust, modify, or alter the content itself*

*Remix – the right to combine the original or revised content with other open content to create something new*

*Redistribute – the right to share copies of the original content, your revisions, or your remixes with others*

Im Zusammenhang mit Versionen spielen vor allem die Aspekte der Nachnutzung (reuse), Änderung (revise) und Vermischung (remix) eine Rolle, indem durch die Bearbeitung von OER neue Versionen erstellt und kollaborativ weiterentwickelt werden können. Plattformen zur Versionsverwaltung, wie GitHub, umfassen mit den Möglichkeiten des Kopierens (retain) und Teilens (redistribute) von Materialien die weiteren Aspekte der 5Rs. Somit stellen OER einen Anwendungsfall für Versionsverwaltung dar.

### 2.1 Prozesse der Versionsverwaltung

Prozesse von Versionsverwaltungssystemen<sup>3</sup>, die in der Entwicklung von Software eingesetzt werden, können auf verschiedene Anwendungsszenarien von Lernressourcen übertragen werden, sodass sich Einsatzszenarien für OER und der kollaborativen Erstellung und Nutzung von Materialien in Lernumgebungen anbieten. Mit der *Fork*-Methode können Projekte kopiert und unabhängig weiterentwickelt werden. Damit können auch Änderungen und individuelle Anpassungen an Lernressourcen vorgenommen werden. Bearbei-

---

<sup>3</sup> Hier am Beispiel von GitHub dargestellt

ter eines Forks können ihre Änderungen mittels eines *Pull Requests* an die Original-Ressource zurückspielen. Der Urheber entscheidet schließlich über die Annahme der Änderungen, die durch einen *Merge* mit dem Original zusammengeführt werden können. Darüber hinaus können Änderungen des Originals automatisch an Forks übermittelt werden, was einen wesentlichen Vorteil gegenüber der Weitergabe einer Datei darstellt. Kooperatives Arbeiten wird ebenso mit *Branches* unterstützt, indem unabhängig voneinander erstellte Elemente in ein Material integriert werden können. Die Speicherung von Änderungen in *Commits* und die Nachvollziehbarkeit durch eine Versionshistorie können darüber hinaus für das gemeinsame Arbeiten an Dokumenten förderlich sein.

## 2.2 Anwendungsfälle der Versionsverwaltung für OER

Bei der Versionsverwaltung von OER lassen sich verschiedene Szenarien unterscheiden, die mit der Anzahl der Ressourcen und der Ersteller bzw. Bearbeiter zusammenhängen. Neue Versionen entstehen bei der Arbeit von einem oder mehreren Urhebern an einem Material, wo vor allem Änderungskommentare und Versionshistorie zur Nachvollziehbarkeit bei der kollaborativen Arbeit hilfreich sein können. Andere Szenarien der Versionsverwaltung beziehen sich auf die Kopie einer Ressource durch einen Nachnutzenden bzw. Bearbeitenden, sodass zwei Ressourcen von unterschiedlichen Personen existieren. Dies kann mit einem Fork gleichgesetzt werden.

Auch in Kursumgebungen sind die kollaborativen Funktionalitäten von GitHub für die Interaktion zwischen Lernenden und Lehrenden einsetzbar. Zwar spiegelt GitHub als Plattform zur Ablage von Kursmaterialien die Funktion von Lernmanagementsystemen wider, geht aber insofern über das Teilen von Ressourcen hinaus, dass sowohl Studierende als auch andere Lehrende die Möglichkeit haben, die Materialien direkt zu nutzen und zu ändern. Durch Bearbeitungen oder Feedback können sie einen Beitrag leisten und werden in die Entwicklung von Materialien einbezogen. [Za15]

## 3 GitHub-Analyse

Um Möglichkeiten der Anwendung der Versionsverwaltung für Open Educational Resources auszuloten, wurde GitHub als Plattform mit aktiver Verwendung im Bereich der Softwareentwicklung ausgewählt. Bei der Analyse sollen die Prozesse der Versionsverwaltung untersucht werden, um der Frage nachzugehen, wie diese beim Umgang mit Bildungsmaterialien genutzt und wie häufig diese Prozesse angewendet werden, um Rückschlüsse auf das Nutzungs- und Interaktionsverhalten im Umgang mit OER zu ziehen.

### 3.1 Vorgehen

Zunächst wird auf das Konzept der GitHub Topics zurückgegriffen, bei dem Repositories von Projektbesitzern mit einem oder mehreren Topics gelabelt werden können, sodass eine

thematische Ausrichtung und inhaltliche Überschneidungen sichtbar werden [Gi20]. In der Analyse wurden 33 Topics ermittelt, die mit Lehr-/Lernmaterialien in Bildungskontexten in Zusammenhang stehen. Diese Topics umfassen Bildungsressourcen zu Kursen, Vorlesungen oder Workshops. So konnten 3325 zugehörige Repositories ermittelt werden. Diese wurden über die GitHub API gesammelt, bei der es sich um eine REST API handelt, die Zugriff aus einem Programm heraus ermöglicht, sodass Daten automatisiert erfasst werden können [Gi20b]. Es ist zu beachten, dass Topics von Projektbeteiligten individuell vergeben werden, sodass weitere Repositories existieren, die die Kriterien eines Bildungsmaterials erfüllen, aber ohne entsprechende Topics nicht Teil der Stichprobe sind.

Um Bildungsmaterialien zu identifizieren, wurden die Repositories aller Topics zunächst einer manuellen Analyse unterzogen. Hierzu wurde auf die README-Files der Repositories zurückgegriffen. Dabei handelt es sich um eine zentrale Informationsquelle auf GitHub, da jeder Ersteller eines Repositories dazu aufgefordert wird in einer autogenerierten Datei sein Projekt näher zu beschreiben. Kriterien bei der Auswahl der Repositories als Bildungsmaterialien umfassen einen thematischen Bildungs- bzw. Lehr/Lernschwerpunkt, Formate, wie Vorlesungen, Workshops, Kurse oder Selbstlernmaterialien sowie Materialarten für Lehr/Lernressourcen, wie Textdokumente, Präsentationsfolien, Lehrbücher oder Notebooks. Dabei konnten 1456 Repositories als Bildungsmaterialien identifiziert werden. Diese wurden näher nach ihrer Anzahl an Forks, Stars, Contributor, Branches, Commits, Issues und Pull Requests analysiert. Diese Daten konnten mit Hilfe eines Python-Skriptes gesammelt werden, welches das Package PyGitHub [Py20] nutzt, um die GitHub API in Python abzubilden sodass GitHub direkt aus einem Python-Skript gecrawlt werden konnte.

### **3.2 Datenauswertung**

Um die Anwendung der Szenarien der Versionsverwaltung für OER zu untersuchen, wurde die Auswertung der Daten zunächst auf die Nachnutzung in Form von Forks, die Erstellung von Versionen mit Commits und die Kooperation von mehreren Personen fokussiert.

Die Anzahl der projektbeteiligten Mitarbeiter geben Hinweise darauf, inwieweit kollaboratives Arbeiten bei der Erstellung von Materialien auf GitHub stattfindet. Während 65% der Repositories von einer Person gepflegt werden, arbeiten bei 30% zwei bis zehn Personen an demselben Projekt. In welchem Umfang Änderungen (Commit) in einem Repository vorgenommen werden, hängt auch mit der Anzahl der Mitarbeitenden zusammen, die an einem Projekt arbeiten (Contributer). So beträgt die durchschnittliche Anzahl der Commits pro Contributor 43,8. Es zeigt sich, dass die Vorteile von GitHub Anwendung finden, da sowohl Materialien von mehreren Personen bearbeitet werden als auch Änderungen an den Materialien vorgenommen und Versionen erstellt werden.

Die Anzahl der Forks pro Repository geben Hinweise auf die Nutzung von Materialien von externen Personen. Abbildung 1 zeigt, dass 593 Repositories, also ca. 40% nie geforkt



wurden. Weitere 42% der Repositories wurden in geringem Umfang, 1- bis 10-mal, geforkt. Nur bei einzelnen beliebten Repositories existiert eine Anzahl von Forks im 1000er Umfang. Diese häufig geforkten Repositories weisen im Vergleich zu den zugehörigen Topics eine inhaltliche Gemeinsamkeit auf. Es handelt sich dabei überwiegend um Materialien, die mit Kursen an Hochschulen oder auf MOOC-Plattformen in Verbindung stehen oder auch Lehrbücher ergänzen.

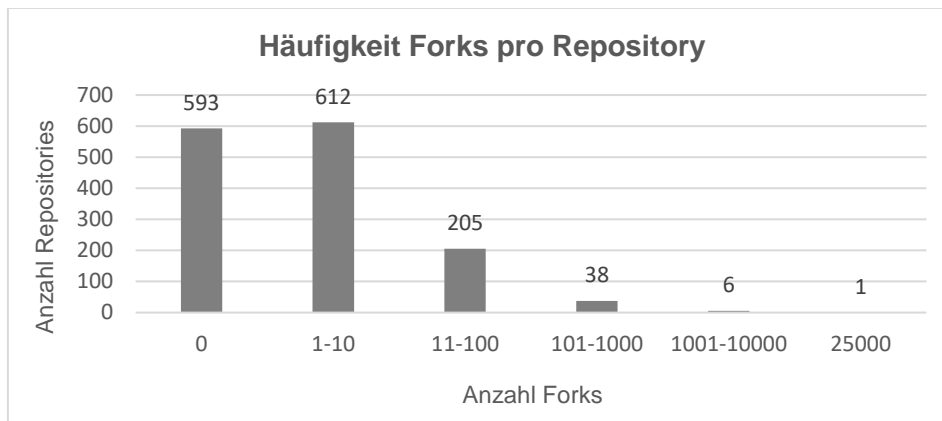


Abb. 1: Häufigkeit von Forks pro Repository

Bei diesen Ergebnissen lassen sich Parallelen zu Softwareprojekten insofern ziehen, dass es viele Projekte mit wenigen Forks und wenige Projekte mit vielen Forks gibt [LRM14]. Hieraus lässt sich allgemein die Schlussfolgerung ableiten, dass bestimmte Materialien besonders häufig von vielen Personen nachgenutzt werden.

#### 4 Zusammenfassung und Ausblick

Die bisherigen Ergebnisse der GitHub-Analyse zeigen, dass die Plattform auch für Bildungsmaterialien verwendet wird, wobei jedoch überwiegend die Möglichkeit der Ablage und Bereitstellung von Ressourcen genutzt wird. Änderungen finden meist von einer Person, dem Projektbesitzer, statt, aber auch in Kooperation von mehreren Personen. Die weiteren Vorteile der Versionsverwaltung, wie Nachnutzung und Bearbeitung von Materialien durch externe Personen, werden in einem geringen Umfang eingesetzt.

Als ein weiterer Schritt bei der Analyse wird angestrebt, die Forks der analysierten Repositories zu untersuchen, um Hinweise auf den Umfang von Änderungen und Weiterentwicklungen sowie die Interaktionen mit dem Original, wie der Zurückführung von Änderungen zu erhalten. Um weitere Repositories zu ermitteln, bei denen es sich um Bildungsmaterialien handelt, muss über das Konzept der Topics hinausgegangen werden. Ein Ansatz besteht darin, die identifizierten Repositories als Grundlage zu nutzen, um einen Clas-

sifiziert nach [Pr18] zu trainieren, damit Repositories automatisch klassifiziert und als Bildungsmaterialien erkannt werden können. Als möglichen Ansatzpunkt für den Algorithmus bieten sich die Repositories an, die mit einer CC-Lizenz versehen sind, um festzustellen, inwieweit hier Bildungsmaterialien enthalten sind.

Die analysierten Bildungsmaterialien weisen einen inhaltlichen Schwerpunkt zu relevanten Themen der Softwareentwicklung auf, was hinsichtlich der Zielgruppe von GitHub nachvollziehbar ist. Daher kommt es bei nicht-technischen Anwendern häufig zu dem Hindernis, dass Funktionalitäten schwer verständlich sind [Za15]. Damit die Potenziale der Versionsverwaltung für die Anwendung von OER in Anlehnung an die 5 R genutzt werden, kann eine Lösungsoption sein, die Prozesse der Versionsverwaltung in eine Nutzeroberfläche zu überführen, die Anwendern außerhalb der Softwareentwicklung den Zugang erleichtert [Ov19]. Dieser Ansatz soll im weiteren Forschungsvorhaben fortgeführt werden.

**Danksagung:** Wir bedanken uns bei Johannes Wendt und Dominik Dyba für ihre Unterstützung bei der Prüfung der Repositories und des GitHub-Crawlings.

## Literaturverzeichnis

- [FF14] Fecher, B.; Friesike, S.: Open Science: One Term, Five Schools of Thoughts. In (Bartling, S.; Friesike, S.): *Opening Science*. Springer, Cham, S. 17-47, 2014.
- [Gi20] GitHub.com: Classifying your repository with topics. <https://help.github.com/en/github/administering-a-repository/classifying-your-repository-with-topics>, Stand: 22.06.2020.
- [Gi20b] GitHub Developer: REST API v3. <https://developer.github.com/v3/>, Stand: 22.06.2020
- [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.
- [LRM14] Lima, A.; Rossi, L.; Musolesi, M.: Coding together at a scale: GitHub as a collaborative social network. In: *Proc. ISWSM*, S. 295-300, 2014.
- [Ov19] Ovia, Steven: Addressing the Technical Challenges of Open Educational Resources. In: *portal: Libraries and the Academy* 19/1, S. 79-93, 2019.
- [Pr19] Prana, G.A.A. et al: Categorizing the Content of GitHub README Files. In: *Empirical Software Engineering*, 24, 1296–1327, 2019.
- [Py20] PyGithub. <https://pygithub.readthedocs.io/en/latest/>, Stand: 22.06.2020
- [Wi14] Wiley, D.: The Access Compromise and the 5th R. In: *Iterating Toward Openness* <http://opencontent.org/blog/archives/3221>, 2014.
- [Za15] Zagalsky, A et al: The Emergence of GitHub as a Collaborative Platform for Education. In: *Proc. CSCW '15*, S. 1906-1917, 2015.

## Corona-Pandemie als Treiber digitaler Hochschullehre

Linda Blömer<sup>1</sup>, Christin Voigt<sup>1</sup> und Uwe Hoppe<sup>1</sup>

**Abstract:** Die Corona-Pandemie beeinflusst die Lehre an Hochschulen weltweit. Bisherige Präsenzveranstaltungen müssen von Lehrenden innerhalb kürzester Zeit in digitale Formate umgewandelt werden. Bereits vor der Corona-Pandemie wurden dafür agile Entwicklungsmethoden und Change Management Konzepte genutzt. Aber können diese Methoden und die dort genannten Aufgaben, die von Lehrenden übernommen werden sollten, auf die aktuelle Situation der Corona-Pandemie übertragen werden? In diesem Beitrag werden auf Grundlage der Literatur Fallbeispiele identifiziert, anhand derer die Aufgaben der Lehrenden sowie bisherige Methoden im Rahmen der agilen Entwicklung digitaler Lehrformate sowie der Gestaltung des digitalen Wandels in der Hochschullehre zusammengetragen werden. Anschließend wird thematisiert, ob und inwieweit Lehrende die Erkenntnisse aus der Theorie auf die aktuelle und zukünftige Praxis anwenden können.

**Keywords:** Agile Entwicklung, Digitale Hochschullehre, Change Management, Corona-Pandemie.

### 1 Einleitung

Die Ausbreitung des Corona-Virus hat im Frühjahr 2020 zu einer Pandemie geführt, die das öffentliche Leben in vielen Ländern maßgeblich beeinflusst. Um die Ausbreitung des Virus einzudämmen, wurde unter anderem die Präsenzlehre an deutschen Hochschulen eingestellt. Dies macht vielerorts die kurzfristige Entwicklung digitaler Lehrformate notwendig. Bisher trieben deutsche Hochschulen digitale Lehre hauptsächlich aus eigenem Antrieb und weniger aufgrund äußerer Rahmenbedingungen voran. [SB18] Doch das Aufkommen der Corona-Pandemie steigert den externen Druck, digitale Lehre an Hochschulen anzubieten. Viele Lehrende stellt dies vor neue Herausforderungen. Bereits vor der Corona-Pandemie wurde über den erfolgreichen Einsatz agiler Methoden zur flexiblen [VKK15] und anpassungsfähigen [Ga16] Entwicklung digitaler Lehrformate berichtet. Auch der Nutzen von Change Management (CM) Prozessen für die effektive Gestaltung des Wandels [He11] ist bereits bekannt. Können diese Erkenntnisse dazu beitragen, die Umstellung auf digitale Lehrformate sowohl während, als auch nach der Corona-Pandemie zu bewerkstelligen? Im Rahmen dieser Arbeit sollen anhand von Fallbeispielen bereits angewandte, agile Methoden, CM Prozesse sowie in diesem Zusammenhang anfallende Aufgaben der Lehrenden identifizieren werden. Das Ziel besteht darin, auf Grundlage der Theorie Handlungsempfehlungen für die Praxis von Lehrenden abzuleiten, die während und auch nach der Corona-Pandemie digitale Lehrkonzepte entwickeln.

---

<sup>1</sup> Universität Osnabrück, IMU, Fachgebiet BWL, Organisation und Wirtschaftsinformatik, Katharinenstraße 1, 49074 Osnabrück, <vorname.nachname>@uni-osnabrueck.de

Im Kern befasst sich diese Arbeit mit den folgenden Forschungsfragen (FF): **FF 1:** Welche Aufgaben der Lehrenden können anhand bisheriger Fallbeispiele im Rahmen der agilen Entwicklung digitaler Lehrformate sowie unter Anwendung von CM Prozessen identifiziert werden? **FF 2:** Inwieweit können Lehrende die Erkenntnisse aus der Theorie auf ihre Praxis während, aber auch nach der Corona-Pandemie anwenden?

Die Arbeit basiert auf zwei systematischen Literaturanalysen, die in Kapitel 2 erläutert werden. Kapitel 3 stellt die Ergebnisse beider Analysen dar (FF 1). Davon ausgehend wird in Kapitel 4 die Ableitung von Handlungsempfehlungen für Lehrende vorgenommen (FF 2). Der Beitrag endet mit einem Fazit, in dem die Grenzen der Arbeit aufgezeigt und ein Ausblick auf zukünftige Forschung gegeben wird.

## 2 Methodisches Vorgehen

Die Grundlage dieses Beitrags bilden zwei systematische Literaturanalysen [WW02], die festgelegten Such- und Auswahlkriterien unterlagen. Zur Identifikation der Quellen dienten Recherchen in den Datenbanken AISnet, Google Scholar, Eric, Science Direct, Scopus und Web of Science. Die Analysen sollen Hinweise liefern, wie eine kurzfristige Umstellung auf digitale Lehrformate gelingen kann - sowohl während, als auch nach der Corona-Pandemie. Denkbar wäre nach der Pandemie beispielsweise die Übertragung entstandener, digitaler Lerninhalte in einen Flipped Classroom (FC) - ein Lehrkonzept, das digitale Lerninhalte als Vorbereitung auf eine Präsenzphase anbietet, in der aktivierende Lernmethoden zum Tragen kommen. [BS12] Um Erkenntnisse des FCs zu berücksichtigen, wurden entsprechende Suchbegriffe in beiden Analysen integriert.

Im Rahmen der ersten Literatanalyse wurde neben den Schlagwörtern zum FC nach dem Stichwort „Change Management“ gesucht. Die Suche wurde auf Veröffentlichungen ab 2015 beschränkt, um möglichst aktuelle Forschungsergebnisse zu erhalten. Insgesamt konnten 294 Quellen identifiziert werden. Nach der Entfernung von 12 Duplikaten wurden die Überschriften, Zusammenfassungen und Stichwörter der verbliebenen Quellen auf ihre Passung überprüft. Quellen, die nicht in einer Fachzeitschrift, auf einer Konferenz, auf Englisch oder Deutsch veröffentlicht wurden, kein CM fokussierten, kein FC oder zumindest Blended Learning (BL) Konzept thematisierten, wurden ausgeschlossen. Dadurch verblieben 44 Quellen für das Lesen der Volltexte, um sie auf ihre Eignung zu überprüfen. Alle Quellen, die CM, FC oder BL nur am Rande behandelten, sich weder auf die Hochschullehre, noch auf Aufgaben der Lehrenden bezogen und kein Fallbeispiel beschrieben, wurden ausgeschlossen. Dies führte zu vier relevanten Quellen.

Die zweite Literatanalyse wurde mit den Schlagwörtern „Agile Process Model“ und „Agile Development“ in Kombination mit den Schlagwörtern zum FC durchgeführt. Dadurch konnten insgesamt 122 Quellen identifiziert werden, unter denen sich sechs Duplikate befanden, die ausgeschlossen wurden. Von den verbliebenen Quellen wurden die Überschriften, Zusammenfassungen und Schlagwörter auf ihre Passung untersucht. Quellen, die keinen Bezug zu agilen Prozessmodellen oder agiler Entwicklung im Zusammenhang mit dem FC, einem BL oder zumindest einem E-Learning Konzept

aufwiesen und Quellen, die nicht auf Englisch oder Deutsch geschrieben waren, wurden ausgeschlossen. Dadurch verblieben 13 Quellen für die Überprüfung auf Eignung durch das Lesen der vollständigen Texte. Auch in diesem Schritt wurde das oben genannte Ausschlusskriterium des fehlenden Bezugs angewandt. Darüber hinaus wurden Quellen ausgeschlossen, die kein Fallbeispiel und keine Aufgaben der Lehrenden beschrieben. Final konnten dadurch drei relevante Quellen identifiziert werden. Die Ergebnisse beider Literaturanalysen werden im Folgenden dargestellt.

### **3 Ergebnisse der Literaturanalysen**

Für die Beantwortung von FF 1 wurden die insgesamt sieben identifizierten Quellen von den Autoren dieser Arbeit strukturiert durchsucht. Dabei mussten die Verantwortlichen und deren Aufgaben häufig aus den Ergebnissen oder Handlungsempfehlungen der Quellen abgeleitet werden. Neben den Lehrenden konnten dadurch weitere Stakeholder wie der Didaktik- und IT-Support sowie das Hochschulmanagement identifiziert werden. Aufgrund der vorliegenden Forschungsausrichtung werden im Weiteren ausschließlich die Stakeholder der Lehrenden betrachtet, zu denen auch ihr Team zählt, das unter anderem aus wissenschaftlichen Mitarbeitern, Projektmitarbeitern und Tutoren bestehen kann.

Von den Quellen, die im Rahmen der ersten Literatanalyse identifiziert wurden, verwendeten zwei keine spezifische CM Methode. [Mo16, Sc16] Während Schoop et al. den Einsatz von E-Learning am Fallbeispiel der TU Dresden beleuchten [Sc16], erläutert Morisse seine Erfahrungen, die er während der Implementierung eines FC, auch als Inverted Classroom bekannt, in zwei Kursen des Fachs „Informatik – Medieninformatik“ an der Hochschule Osnabrück gesammelt hat. [Mo16] Van Twembeke und Goeman interviewten elf Lehrende und stellten ein selbst entwickeltes, konzeptionelles Modell vor, in dem CM Aspekte mit motivationalen Faktoren verknüpft werden, um die Akzeptanz der Lehrenden bezüglich technologiegestütztem Lernen zu erhöhen. [VG18] Collyer und Campbell nutzten ebenfalls die Methode des qualitativen Interviews, verwendeten jedoch das CM Modell von Kotter, um bewährte Vorgehensweisen zur Implementierung technologiegestützten Lernens festzuhalten. [CC15]

Im Rahmen der zweiten Literatanalyse konnten insgesamt drei Fallstudien gefunden werden, die sich mit agilen Methoden des Wandels beschäftigen. [Ga16, MS14, VKK15] Bei Gale et al. war - ähnlich wie während der aktuellen Corona-Pandemie - eine schnelle, agile Entwicklung digitaler Lehrformate aufgrund der Ebola-Epidemie notwendig, um das Gesundheitspersonal in Afrika trotz zeitlicher, örtlicher und finanzieller Beschränkungen bestmöglich weiterbilden zu können. [Ga16] Vogel et al. orientierten sich bei ihrer agilen FC-Entwicklung eines Programmier- und Softwareentwicklungskurses an den agilen Sprints innerhalb von Scrum. [SS13, VKK15] Einer ihrer Kurse wurde darüber hinaus in Anlehnung an das Just-in-Time-Teaching (JiTT) durchgeführt. Zudem haben Vogel et al. den Studierenden Online-Materialien und Fragen zur Verfügung gestellt. [VKK15] JiTT wurde des Weiteren von den Autoren Meissner und Stenger verwendet [MS14], die das

Fach „Elektrotechnik“ an der TH Nürnberg unterrichteten. Mithilfe von JiTT konnten sie ihren Teilnehmern einen lernerzentrierten und individuellen Lernprozess bieten. [MS14]

Alle Aufgaben der Lehrenden (+Team), die im Rahmen der genannten Fallstudien identifiziert werden konnten, werden in Tabelle 1 zusammengefasst dargestellt.

<i>Nr.</i>	<i>Aufgaben während des (agilen) Wandels</i>	<i>Quelle(n)</i>
1	Experten akquirieren	[Ga16, VKK15]
2	Lernziele definieren	[MS14]
3	Bedürfnisse und Anforderungen erfassen	[Ga16, VKK15]
4	(Teil-) Aufgaben definieren	[Ga16, VKK15]
5	Verantwortlichkeiten definieren	[VKK15]
6	Software auswählen	[Ga16, MS14, VKK15, CC15]
7	Räumliche, zeitliche und finanzielle Einschränkungen beachten	[Ga16]
8	(Agile) E-Learning-Strategie entwickeln	[Sc16, Ga16]
9	Didaktische Methoden abwägen	[MS14]
10	Heterogenes Team bilden und zusammenbringen	[Ga16, VKK15, Sc16, VG18]
11	Kommunikationstool für Beteiligte auswählen	[Ga16, VKK15]
12	Kernelemente kommunizieren	[Ga16]
13	Effizientes Arbeiten sicherstellen	[Ga16, VKK15]
14	Zwischenprodukt herstellen	[VKK15]
15	Software entwickeln	[Ga16]
16	Lernplattform aufbauen und testen	[Ga16]
17	Übungsaufgaben ausarbeiten	[MS14]
18	Präsenzphase gestalten	[MS14]
19	Vorhaben an Studierende kommunizieren	[Mo16, VG18]
20	Vorteile von E-Learning vermitteln	[CC15]
21	Studierende zum Selbststudium anleiten	[Mo16]
22	Studierenden Anreize bieten	[VG18]
23	Aus Fehlern lernen	[VKK15, VG18]

Tab. 1: Ermittelte Aufgaben der Lehrenden während des (agilen) Wandels

#### 4 Handlungsempfehlungen für Lehrende

Insbesondere zu Zeiten der Corona-Pandemie steht der schnelle und effiziente Wandel bisheriger, analoger Lehrformate im Vordergrund. Lehrende können agile Methoden wie Scrum dazu nutzen, um in kürzester Zeit digitale Lehrformate herzustellen. Um die Scrum-Prozesse noch zeitsparender zu gestalten, kann die Dauer zwischen den Schleifen verkürzt und somit die Einbindung des Entwicklungsteams beschleunigt werden. Ähnlich wie während der Ebola-Epidemie [Ga16] sollte die Generierung kleiner, realisierbarer Zwischenziele fokussiert werden. Zudem können CM Prozesse den erfolgreichen Wandel unterstützen und die nachhaltige Implementierung digitaler Lehrformate sicherstellen. Gemäß Kotter könnten Fehler im Rahmen eines CM Prozesses dem Wandel schaden und

sollten vermieden werden. Deshalb empfiehlt er, jeden CM Schritt ausreichend zu behandeln, bevor der nächste begonnen wird. [Ko96] Es ist jedoch fraglich, inwieweit dies während der Corona-Pandemie berücksichtigt werden kann. Stattdessen kann das CM Modell von Kotter als übergeordneter Leitfaden gesehen werden, dessen Schritte zwar alle bedacht, bei Bedarf jedoch in unterschiedlicher Reihenfolge durchgeführt werden können. Die in Tabelle 1 gezeigten Aufgaben können von Lehrenden (+Team) als Checkliste herangezogen werden. Aufgrund individueller Rahmenbedingungen und zeitlicher Kapazitäten sollte jedoch jede/r Lehrende eigene Prioritäten bei der Auswahl und Reihenfolge der Aufgaben setzen. Zu den zentralen, in Tabelle 1 aufgelisteten und in mindestens drei der Quellen genannten Aufgaben zählt die Auswahl geeigneter Software und die Bildung eines heterogenen Teams. Einen wichtigen Bestandteil des agilen Prozesses stellt zudem die Ermittlung der Bedürfnisse und Anforderungen, der Aufbau eines Kommunikationstools sowie die Definition der (Teil-) Aufgaben dar. Des Weiteren ist die Evaluation der unter Hochdruck entstandenen E-Learning Materialien ratsam. Insgesamt empfiehlt es sich, die Digitalisierung der eigenen Lehre während der Corona-Pandemie nicht nur als Übergangslösung, sondern als mögliche Bereicherung der eigenen, zukünftigen Lehre zu betrachten. Denn die während der Pandemie entwickelten, digitalen Lerninhalte können Lehrende auch nach der Pandemie, beispielsweise in Form eines FC-Konzepts, nutzbringend verwenden.

## 5 Fazit

In dieser Arbeit konnten verschiedene Modelle sowie insgesamt 23 spezifische Aufgaben der Lehrenden und ihrer Teams anhand bisheriger Fallbeispiele aus dem CM und der agilen Entwicklung digitaler Lehrformate identifiziert (**FF1**) und in Anlehnung an die Ergebnisse Handlungsempfehlungen abgeleitet werden (**FF2**). Die gewonnenen Erkenntnisse sollen Lehrende sowohl während der außergewöhnlichen Situation aufgrund der Corona-Pandemie, als auch zukünftig im Rahmen der Entwicklung digitaler Lehrformate unterstützen. Die Aufgaben und Handlungsempfehlungen können insbesondere Lehrenden, die bisher wenige Kenntnisse bezüglich der Transformation digitaler Lehre gesammelt haben, als Orientierungshilfe für die Umsetzung dienen.

Die Grenzen dieser Arbeit resultieren zum einen aus den Beschränkungen der Literaturanalysen auf den FC als eine Form von E-Learning sowie auf fallstudienbezogene Forschung. Zukünftig ist die Evaluation der Aufgaben durch eine Befragung der Lehrenden geplant. Dabei soll sowohl deren Relevanz, die Relevanz bisher nicht genannter Aufgaben sowie die Zuweisung von Verantwortung überprüft werden. Zudem gilt es weitere Stakeholder und deren Zusammenarbeit im Rahmen des digitalen Wandels zu berücksichtigen. Darüber hinaus sollte erforscht werden, inwiefern die Corona-Pandemie den digitalen Wandel beeinflusst (hat) und ob die ausgelöste Dynamik digitaler Hochschullehre in der Lage ist, die Pandemie langfristig zu überdauern.

**Literaturverzeichnis**

- [BS12] Bergman, J.; Sams, A.: Flip your classroom: Reach every student in every class every day. Alexandria : International Society for technology in Education, 2012.
- [CC15] Collyer, S.; Campbell, C.: Enabling Pervasive Change: A Higher Education Case Study. In: EdMedia+ Innovate Learning: Association for the Advancement of Computing in Education (AACE), S. 249-255, 2015.
- [Ga16] Gale, T.; Chatterjee, A.; Mellor, N.; Allan, R.: Health worker focused distributed simulation for improving capability of health systems in Liberia. In: Simulation in Healthcare Bd. 11, Nr. 2, S. 75-81, 2016.
- [He11] Herzfeldt, A.; Kristekova, Z.; Schermann, M.; Krcmar, H.: A Conceptual Framework of Requirements for the Development of E-Learning Offerings from a Product Service System Perspective. In: Americas Conference on Information Systems (AMCIS), S. 1-9, 2011.
- [Ko96] Kotter, J.: Leading Change: Why Transformation Efforts Fail. In: Harvard Business Review, S. 57-68, 1995.
- [MS14] Meissner, B.; Stenger, H.-J.: Agiles Lernen mit Just-in-Time-Teaching. In: Teaching Trends 2014, S. 121-136, 2014.
- [Mo16] Morisse, K.: Inverted Classroom in der Hochschullehre – Chancen, Hemmnisse und Erfolgsfaktoren. In: Das Inverted Classroom Modell. Begleitband zur 5. Konferenz Inverted Classroom and Beyond, S. 1-11, 2016.
- [No11] Novak, G.: Just-in-time teaching. In: New Directions for Teaching and Learning Bd. 2011, Nr. 128, S. 63-73, 2011.
- [SS13] Schwaber, K., Sutherland, J.: The scrum guide-the definitive guide to scrum: The rules of the game, 2013.
- [Sc16] Schoop, E.; Köhler, T.; Börner, C.; Schulz, J.: Consolidating eLearning in a Higher Education Institution: An Organisational Issue integrating Didactics, Technology, and People by the Means of an eLearning Strategy, S. 39-50, 2016.
- [SB18] Schünemann, I.; Budde, J.: Hochschulstrategien für die Lehre im digitalen Zeitalter: Keine Strategie wie jede andere!. In: Arbeitspapier 38, Hochschulforum Digitalisierung, 2018.
- [VG18] Van Twembeke, E.; Goeman, K.: Motivation gets you going and habit gets you there. In: Educational Research Bd. 60, Nr. 1, S. 62-79, 2018.
- [VKK15] Vogel, B.; Kilamo, T.; Kurti, A.: Teaching distributed agile development to software professionals: a flexible approach. In: Proceedings of the 2015 European Conference on Software Architecture Workshops: ACM, S. 1-8, 2015.
- [WW02] Webster, J.; Watson, R.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly Bd. 26, Nr. 2, S. xiii–xxiii, 2002.



## **Poster**



## VeraSQL: An educational client for relational databases

André Greubel<sup>1</sup>, Theresa Rudolph<sup>1</sup> and Martin Hennecke<sup>1</sup>

**Abstract:** In this paper, we present VeraSQL, an educational database client specifically designed to support the transition between spreadsheets and relational databases as taught in many German school curricula. Most importantly, VeraSQL can execute SQL queries on CSV-files. VeraSQL is written in Java, open source, and can be used and downloaded free at [www.verasql.de](http://www.verasql.de).

**Keywords:** SQL, query languages, educational database client, database curriculum

### 1 Overview

In German school curricula, lessons on databases and SQL follow lessons on spreadsheets and corresponding software like Microsoft Excel or Libre Office Calc. While learning about these topics, learners often contact ideas and methods, which can also be utilized for databases. However, the current set of professional and educational tools commonly used in classrooms neither supports this transition nor highlights similarities between these topics.

Educational software usually only offers an *educational SQL environment*, rather than a full *database client*. Most importantly, it is not easily possible to select a custom data source (e.g., data from spreadsheets) in most environments. For example, it is not possible to change the data source in BlocklySQL [Pö19] at all. Other solutions like DBSnap++ [Si] can import a spreadsheet as data source, but are unable to easily build and execute SQL-queries. Due to this lack of suitable educational software, some teachers and schoolbooks (e.g., [Br07]) even opt to use professional database management software instead.

Our educational tool, VeraSQL<sup>2</sup>, was explicitly designed to close this gap by:

- Offering a visual user interface for SQL-queries. This interface, like block-based environments, enables the creation of queries with drag-and-drop.
- Offering a custom SQL-compiler able to execute a given SQL-query step-by-step on CSV-spreadsheets.

---

<sup>1</sup> University of Wuerzburg, Institute for Computer Science, Mathematik West (30), 97074 Wuerzburg  
{andre.greubel, martin.hennecke}@uni-wuerzburg.de, theresa.rudolph@stud-mail.uni-wuerzburg.de

<sup>2</sup> Vera is a German Acronym for “visual creation of queries” (german: [V]isuelle [Er]stellung von [A]bfragen)

- Simplifying presentation of exercises by showing a given query simultaneously with its result.

## 2 VeraSQL

In VeraSQL, queries are displayed on the left and can be executed and stored to / loaded from an SQL-file with the left buttons in the bottom control area. Clicking on the shape of an expression in the bottom control area adds a condition to the end of the corresponding expression in the query display area on the left.

The right buttons in the QueryControlArea allow for the selection of a data source. VeraSQL can either be connected to a database server or use local data sources.

The top button connects to a database server. In this case, queries are sent to the server and the result is displayed in the MainArea.

The lowest button in the right row loads a single CSV-file as a table into VeraSQL. Similarly, the middle button loads every CSV-file in a directory as a table into VeraSQL. In these cases, VeraSQL automatically enters *local database mode*. In this mode, a custom SQL-compiler is used to locally execute the query on the selected spreadsheet-tables (instead of sending it to a database server).

This local database mode is one of the key features of VeraSQL and currently responsible for roughly 40% of the 13.000 nonempty lines of Java-Code.


The local compiler responsible for the execution of the SQL-query is based on a simplified custom grammar in ANTLR. Most of the time, the compiler follows SQL-compliant behaviour. However, as CSV-files do not specify data formats, VeraSQL automatically infers them. If every entry in a column can be parsed to a `BigDecimal`, the column is represented that way. Otherwise, the column type defaults to a `java.lang.String`. Because of this, we recommend formatting dates and times using ISO 8601 in the user data to prevent non-standard compliant sorting.

A more detailed technical description of VeraSQL (including limitations and technical details about the compiler), as well as VeraSQL itself, is available at [www.verasql.de](http://www.verasql.de).

## Bibliography

- [Br07] Brichzin, P. et al.: Informatik. Oldenbourg, München, 2007.
- [Pö19] Pöhner, N. et al.: BlocklySQL: A new block-based editor for SQL: Proceedings of the 14th Workshop in Primary and Secondary Computing Education, 2019; pp. 1–2.
- [Si18] Silva, Yasin N. et al.: DBSnap++: Proceedings of 23rd Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE'18); pp. 170–175.

## 360-degree Models in Environmental Engineering Education: an Explorative Case Study

Mario Wolf <sup>1</sup>, Heinrich Söbke<sup>1</sup>, Florian Wehking<sup>1</sup>, Stefanie Hörnlein<sup>1</sup>

**Abstract:** 360-degree models are defined as the algorithmic merging of multiple 360-degree images of an enclosed space creating a single visual model of the space that can be viewed from various positions within the space. As means of virtual reality, 360-degree models can be understood as a medium with impact on learning. Aiming at evaluating the educational potential of 360-degree models in the field of environmental engineering, an explorative pilot study featuring an innovative sanitation system has been conducted (N=17). Quantitative measurements of motivation, emotion and usability show positive results. Guided interviews revealed both key points and problems to be addressed. Although the learning efficiency is still to be validated, the study renders 360-degree models as promising learning tools that might enrich tertiary education in environmental engineering.

**Keywords:** Virtual Reality, 360-degree Model, Education, Environmental Engineering, User Experience, Motivation, Emotion.


### 1 An Explorative Case Study

360-degree models are algorithmically assembled 360-degree images of objects, for example spaces, which allow observers to view the objects almost from all directions. 360-degree models differ from 3D models generated by 3D scanners in lacking quantified depth information [SK14]. An important field of application of 360-degree models, which has been established for several years [Am10], is virtual visits to real estate, which allow prospective buyers or tenants to gain an impression of the property independent of time, location and people [Fe18]. Beyond its use for commercial purposes, 360-degree models are also suitable as a learning tool, for example [Ph18] describe a 360-degree model teaching safety on construction sites.

In environmental engineering, locations that can be depicted using 360-degree-technology include not only structures, but also laboratories and test facilities. Thus, the aim of the explorative pilot study is exploring the potential of 360-degree models as learning tools. The 360-degree model of an innovative sanitation system, having been augmented by textual, graphical and video annotations and aiming at public awareness rising is explored by environmental engineering students regarding usability, motivation and emotions. The

---

<sup>1</sup> Bauhaus-Universität Weimar, Bauhaus-Institute for Infrastructure Solutions (b.is), Coudraystraße 7, 99423 Weimar, {ulrich.mario.wolf|heinrich.soebke|florian.wehking|stefanie.hoernlein}@uni-weimar.de

 <https://orcid.org/0000-0002-0972-1117>

study comprised of four phases: **Exploration** consisted Ten minutes lasting exploration phase of the 360-degree model either on a tablet (N=8) or on a desktop computer (N=9). A **questionnaire** asked besides prior knowledge, demographics and a self-estimation of the learning outcomes, for usability [MHN12], emotions [Pe11] and motivation [RVB01]. A **post test** comprising of five multiple choice questions aimed at evaluating the knowledge gained from the exploration of the 360-degree model. **Semi-structured interviews** of 5 minutes duration gathered further feedback. Table 1 describes identified strengths and weaknesses.

Strengths	Weaknesses
<ul style="list-style-type: none"> <li>• Provides holistic overview over topic</li> <li>• Information are highly understandable and memorable</li> <li>• Useful support to create mental models</li> <li>• Opportunity to present new projects quickly</li> <li>• High motivation of usage</li> </ul>	<ul style="list-style-type: none"> <li>• Uncertainty about kind and extend of information</li> <li>• Deficient structure of information</li> <li>• Complex approach to learn detailed information</li> <li>• Specific information searched are hard to find</li> </ul>

Table 1 Educational use of a 360-degree model: summarized results

## Bibliography

- [Fe18] Felli, F.; Liu, C.; Ullah, F.; Sepasgozar, S. M. E.: Implementation of 360 videos and mobile laser measurement technologies for immersive visualisation of real estate & properties. In *Education Building Professionals for the Future in the Globalised World*, 2018, 2; S. 294–305.
- [MHN12] Müller, J.; Heidig, S.; Niegemann, H. M.: Evoking emotional dimensions in HCI - development of the questionnaire user experience (QUX), 2012.
- [Pe11] Pekrun, R.; Goetz, T.; Frenzel, A. C.; Barchfeld, P.; Perry, R. P.: Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). In *Contemporary Educational Psychology*, 2011, 36; 36-48.
- [Ph18] Pham, H. C.; Dao, N. N.; Pedro, A.; Le, Q. T.: Virtual field trip for mobile construction safety education using 360-degree panoramic virtual reality. In *The International journal of engineering education*, 2018; S. 1174–1191.
- [RVB01] Rheinberg, F.; Vollmeyer, R.; Burns, B. D.: QCM. A questionnaire to assess current motivation in learning situations. In *Diagnostica*, 2001, 47; 57–66.
- [SK14] Straub, J.; Kerlin, S.: Development of a Large, Low-Cost, Instant 3D Scanner. In *Technologies*, 2014, 2; S. 76–95.

## Adapting virtual trainings of applied skills to cognitive processes in medical and health care education within the DiViFaG project

Carolin Hainke<sup>1</sup>, Thies Pfeiffer<sup>2</sup>

**Abstract:** The use of virtual reality technology in education rises in popularity, especially in professions that include the training of practical skills. By offering the possibility to repeatedly practice and apply skills in controllable environments, VR training can help to improve the education process. The training simulations that are going to be developed within this project will make use of the high controllability by evaluating behavioral data as well as gaze-based data during the training process. This analysis can reveal insights in the user's mental states and offers the opportunity of autonomous training adaption.

**Keywords:** virtual reality; training simulation; medicine; health care; cognitive load; eye-tracking

### 1 Research Context

Since virtual reality (VR) became more affordable over the last years, awareness of its potential for educational purposes has risen. It has been shown that the use of VR techniques can help to promote skill acquisition during training. Additionally, the VR training environment is highly controllable and can be adjusted to the learner's skill levels [CER19]. By making use of those benefits, VR training simulations can improve education as in the context of medicine and health care, since those educations include the training of practical skills. During their education students often state lacking the possibility of training practical skills sufficiently [Fe04]. To further enhance the training, it is important to reduce the amount of extraneous cognitive load while keeping the level of difficulty of the training context cognitively demanding [Sw11].

### 2 Research Question

During the project multiple VR training scenarios will be developed by all project partners to make use of the mentioned benefits of VR training for medical and health care education. By integrating the scenarios into existing learning management systems, students can easily access them and teachers can adjust them to their personal needs.

---

<sup>1</sup> University of Applied Science Emden/ Leer, Faculty of Technology, Constantiaplatz 4, 26721 Emden, carolin.hainke@hs-emden-leer.de

<sup>2</sup> University of Applied Science Emden/ Leer, Faculty of Technology, Constantiaplatz 4, 26721 Emden, thies.pfeiffer@hs-emden-leer.de

To enable an even more autonomous and individual training, the scenarios will be extended by online user data evaluation, including behavioral data as well as gaze-based data. This enables the measurement of mental states such as cognitive load by, for example, analyzing pupil oscillation [Du20] or the task-dependent pupil size [HP17]. The training environment then adapts to those measurements by offering more hints or guidance for more insecure or novice users or changing the task design from a recall to a recognition task to prevent cognitive overload (Fig.1). The additional feedback, as well as changing the task for more advanced learners to a design that builds on prior knowledge, are based on the cognitivist learning theory [BW77]. The training adaptations will be evaluated to analyze the impact of the considered changes regarding improved learning outcome, higher motivation and impact on the long-term memory.

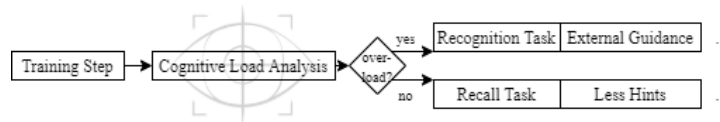


Fig. 1: Possible decision process of the application based on user data analysis

The project has started in January 2020. Until September we will have implemented three scenarios that cover interaction techniques such as eye- and hand-tracking as well as speech recognition. We look forward to demonstrating examples at the conference. DiViFaG is sponsored by the Bundesministerium für Bildung und Forschung.

## Bibliography

- [BW77] Bandura, A.; Walters, R. H.: Social learning theory (Vol. 1). Englewood Cliffs, NJ: Prentice-hall, 1977.
- [CER19] Concannon, Brendan J.; Esmail, Shaniff; Roduta Roberts, Mary: Head-Mounted Display Virtual Reality in Post-secondary Education and Skill Training. *Frontiers in Education*,4:80, 2019.
- [Du20] Duchowski, A. T.; Krejtz, K.; Gehrer, N. A.; Bafna, T.; Bækgaard, P.:The Low/High Index of Pupillary Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12), 2020.
- [Fe04] Federkeil, G. , *Ergebnisse einer vergleichenden Absolventenbefragung Humanmedizin des Centrums für Hochschulentwicklung*, 2004.
- [HP17] Hainke, C.; Pfeiffer, T.: Messen mentaler Auslastung in einer VR-Umgebung basierend auf Eyetrackingdaten. In *Virtuelle und Erweiterte Realität-14. Workshop der GI-Fachgruppe VR/AR*, 2017.
- [Sw11] Sweller, John: Cognitive load theory. In: *Psychology of learning and motivation*, volume 55,pp. 37–76. Elsevier, 2011.



## Selbsteinschätzung leicht gemacht

### Studienergebnisse aus der Theoretischen Informatik

Arno Wilhelm-Weidner<sup>1</sup>

**Abstract:** Der Posterbeitrag fasst Ergebnisse von fünf Studien zusammen, in denen eine einfache Form der Selbsteinschätzung von Bachelor-Studierenden in der Theoretischen Informatik für mehrere Themengebiete genutzt wurde. Die Ergebnisse übertreffen bisherige Angaben zur Selbsteinschätzung aus der Literatur.

**Keywords:** Theoretische Informatik, Selbsteinschätzung, e-Learning

## 1 Einleitung

Den eigenen Kenntnisstand und die Kompetenzen einschätzen zu können ist ein relevanter Bestandteil des Lernprozesses. Wenn Selbsteinschätzungen von Lernenden kurzzeitig erfolgen und angemessen zutreffend sind, können sie Lehrenden helfen, ihre Lehre an wesentlichen Stellen anzupassen. Selbsteinschätzungen werden dabei in der Literatur allerdings häufig in umfangreicherer Form eingesetzt, als es in den Studien der Fall war, die der Posterbeitrag beschreibt.

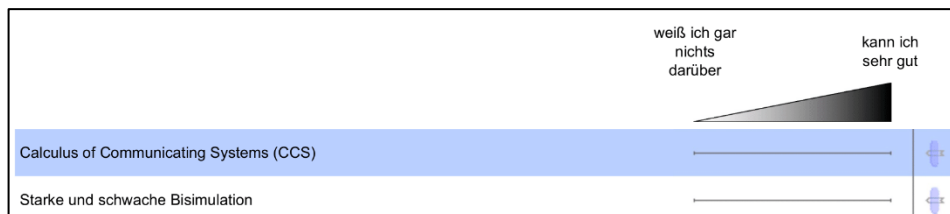


Abb. 1: Beispiel für die Selbsteinschätzung.

Ein Großteil der Ergebnisse des Beitrags entstand im Rahmen von Studien zu Lerneinheiten zur Theoretischen Informatik im Lernmanagementsystem Moodle. Thematisch wurden zwei Module (zu formalen Sprachen und reaktiven Systemen) mit je zwei Lerneinheiten an verschiedenen Standorten genutzt. Die gesamte Evaluation ist in [Wi20] zu finden. Im Rahmen der Evaluation wurde in sechs verschiedenen Modulen an fünf Standorten in Umfragen untersucht, inwiefern sich die Nutzung der Lerneinheiten auf Motivation und Kompetenzerwerb auswirkt. Neben Fragen zum Kompetenzerwerb waren

<sup>1</sup> Technische Universität Berlin, FG Modelle und Theorie Verteilter Systeme, Ernst-Reuter-Platz 7, 10587 Berlin, arno.wilhelm-weidner@tu-berlin.de

auch Selbsteinschätzungen enthalten. Diese erfolgten im Rahmen der Umfragen durch Angabe eines Themas und Einschätzung je über einen Regler auf einer Skala von „weiß ich gar nichts darüber“ bis zu „kann ich sehr gut“ (siehe Abbildung 1). Um Einflussfaktoren herauszuarbeiten wurden mittels Kendalls  $\tau$  unter anderem Korrelationskoeffizienten zwischen den Kompetenzen und den entsprechenden Selbsteinschätzungen für die Studien an der TU Berlin, der Uni Potsdam und der Uni Duisburg-Essen berechnet (siehe Tabelle 1). Die Kompetenz-Fragen orientierten sich dabei an Modulhandbüchern vergleichbarer Module und bestehenden Prüfungsfragen.

Thema	Standort	Umfrage	n	Kendalls $\tau$	p-Wert
FoSA	Berlin	2	45	0,338	0,003
		3	36	0,287	0,026
ReSyst	Berlin	2	29	0,127	0,368
		3	26	0,378	0,015
FoSA	Potsdam	2	14	0,553	0,009
		3	9	0,458	0,114
FoSA	Duisburg	2	6	0,276	0,444
		3	4	0,548	0,279

Tab. 1: Korrelationsergebnisse der ersten vier Studien.

Da mehrere statistisch interessante Ergebnisse auftauchten, bei denen Studierende mit einer hohen Selbsteinschätzung auch gute Ergebnisse in den passenden Kompetenzfragen aufwiesen, wurde im Januar 2020 im Modul zu formalen Sprachen an der TU Berlin eine weitere Studie durchgeführt. Die Umfrage enthielt sechs Fragen zu Potenzmengen und eine Selbsteinschätzung wie in Abbildung 1, unter anderem mit dem Punkt „Potenzmengen“. Für die Teilnehmer ( $n = 42$ ) ergab die Korrelation aus Kompetenzen und Selbsteinschätzung einen Wert von  $\tau = 0,329$  und  $p = 0,006$ . Der vorher mehrfach auftretende Zusammenhang, dass Studierende mit einer hohen Selbsteinschätzung auch gute Kompetenzergebnisse haben, hat sich bestätigt. Die Ergebnisse deuten auf einen Zusammenhang zwischen Kompetenzerwerb und dieser einfachen Form der Selbsteinschätzung hin. Allerdings unterliegen die Ergebnisse aufgrund der einfachen Natur der Selbsteinschätzungen Einschränkungen.

## Literaturverzeichnis

- [Wi20] Wilhelm-Weidner, A.: Conception and Evaluation of E-Learning Units Regarding Motivation and Acquired Competencies for Theoretical Computer Science at University Level. Dissertation. Technische Universität Berlin, 2020.

## Immersive Virtual Reality Training for the Operation of Chemical Reactors

Yusra Tehreem<sup>1</sup>, Thies Pfeiffer<sup>2</sup>

**Abstract:** This paper discusses virtual reality (VR) training for chemical operators on hazardous or costly operations of chemical plants. To this end, a prototypical training scenario is developed which will be deployed to industrial partners and evaluated regarding efficiency and effectiveness. In this paper, the current version of the prototype is presented, that allows life-sized trainings in a virtual simulation of a chemical reactor. Building up on this prototype scenario, means for measuring performance, providing feedback, and guiding users through VR-based trainings are explored and evaluated, targeting at an optimized transfer of knowledge from virtual to real world. This work is embedded in the Marie-Sklódowska-Curie Innovative Training Network CHARMING<sup>3</sup>, in which 15 PhD candidates from six European countries are cooperating.

**Keywords:** Virtual Reality; Training; Chemical Reactor; Safety

### 1 Introduction

In chemical plants, mistakes during training can cause safety, time, and cost issues [Na12]. Besides, there is also a hesitation from a trainee to explore the scenarios to their worst and best outputs. Thus, virtual reality is chosen as an immersive solution because it can provide safe training options for hazardous scenarios [NKS14]. Our research is under CHARMING<sup>3</sup>, which is focusing on training of children, students and employees in chemistry, chemical engineering and chemical industry. Our focus is to develop a VR training prototype and evaluate it with the employees of industrial partners of the project. In the section below, the current state of prototype and training plan is discussed.

### 2 Virtual Reality-based Training

A VR-based prototype is developed which allow users to perform a chemical procedure of Butyllithium. From requirements analysis, this reaction is very dangerous and relevant to the products formed in chemical plants. Besides, a 3D model of a full chemical reactor is embedded which is required for this reaction and for multiple chemical reactions in industry making it a valid model for training. According to the current prototype, users

---

<sup>1</sup> HS Emden/Leer, FB Technik/E+I, Constantiaplatz 4, Emden, 26723, yusra.tehreem@hs-emden-leer.de

<sup>2</sup> HS Emden/Leer, FB Technik/E+I, Constantiaplatz 4, Emden, 26723, thies.pfeiffer@hs-emden-leer.de

<sup>3</sup> This project has received funding from the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under Grant Agreement No 812716.

can perform the procedure by controlling the virtual reactor both manually and via a simulated computer screen inside the VR environment. This is to copy the original procedure by learning both automated and manual controls of today's chemical plants.



Fig. 1: The VR Prototype would span three levels (left) of a large building in real life, which would be costly to realize as a pure training environment. The main area (middle) hosts the reactor and the control panel (right, enlarged for visibility reasons)

The next step is to induce multiple training modes accompanied by feedback and reporting strategies. For employees, it can be a first VR attempt. Thus, rather than confronting them with all knowledge of VR and chemical operations at once, a step by step knowledge acquisition will be provided in the form of training modes (as in figure below). This will enable exploration, practicing with guidance, continuous feedback and then performance evaluation for both trainee and the trainer. The prototype will be assessed to industrial partners and it will be tested against their limitations of classroom lectures and pilot plants. The target is to achieve a VR training design that fills up the real weaknesses of traditional training and complements the whole training system.

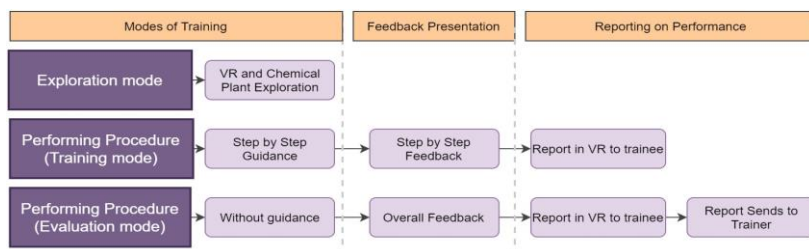


Fig. 2: Training Layout in Virtual Reality

## Bibliography

- [Na12] Nazir, Salman; Totaro, Roberto; Brambilla, Sara; Colombo, Simone; Manca, Davide: Virtual Reality and Augmented-Virtual Reality as Tools to Train Industrial Operators. Computer Aided Chemical Engineering 22nd European Symposium on Computer Aided Process Engineering, p. 1397–1401, 2012.
- [NKS14] Nakai, Atsuko; Kaihata, Yuta; Suzuki, Kazuhiko: The Experience-Based Safety Training System Using VR Technology for Chemical Plant. International Journal of Advanced Computer Science and Applications, 5(11), 2014.

## LAYA will in die Schule

### Eine Anforderungsanalyse für den inklusiven, kollaborativen Einsatz einer Lernsoftware in der Sekundarstufe

Sebastian Claus<sup>1</sup>, Niels Pinkwart<sup>2</sup>

**Abstract:** LAYA (Learn as you are!) wird bisher im Kontext der Erwachsenenbildung eingesetzt. Ziel dieses Posters ist es die Anforderungsanalyse für einen inklusiven, kollaborativen Einsatz in der Sekundarschule mit ersten Erkenntnissen darzustellen. Dem Tagungsmotto „Educational Realities“ wird durch dem gewählten Participatory Design-Ansatz und seinem Bezug zur Grounded Theory Methodology besonders Rechnung getragen.

**Keywords:** Inklusion, Computer-supported Collaborative Learning, Participatory Design, Requirements Engineering

## 1 Einleitung

Die inklusive Lernplattform LAYA wird bisher in der Erwachsenenbildung eingesetzt. Patzer et al. diskutieren Webvideoplayer für die Videokomponente [PP19] von LAYA. Der vorliegende Beitrag fasst Inklusion weit, d. h., dass über Behinderung hinausgehend auch soziale und kulturelle Dimensionen Berücksichtigung finden. Eine Recherche zeigt, dass lediglich eine Abschlussarbeit von Naoufal [NKN19] sich mit Peer-Interaktionen in einem inklusiven E-Learning-Szenario befasst. Peer-Interaktionen und -Kooperationen in inklusiven Settings sind also ein weißer Fleck in der E-Learning-Forschung. Dieser Beitrag stellt erste Anforderungsideen vor, die aus Lehrkräfteinterviews hervorgehen. Die Postervorstellung kann diese, im Rahmen der fortgeschrittenen Datenerhebung und -auswertung, kritisch reflektieren.

## 2 Methodisches Vorgehen

Zur Anforderungsanalyse wird ein Participatory Design-Ansatz verfolgt, bei dem die Schüler\*innen- (aktuell sechs) und Lehrer\*innenperspektiven (zwei) durch teilstrukturierte Interviews berücksichtigt werden. Aufgrund der momentanen COVID-19-Pandemie können lediglich die Lehrkräfteinterviews online fortgeführt werden.

---

<sup>1</sup> Humboldt-Universität zu Berlin, Institut für Informatik, 10099 Berlin, clausseb@hu-berlin.de

<sup>2</sup> Humboldt-Universität zu Berlin, Institut für Informatik, 10099 Berlin, pinkwart@hu-berlin.de

Die Wahl von Fachlehrkräften als Experten/-innen im Feld und als Praxisakteure/-innen liegt nahe. Die Daten werden um die Schüler\*innenperspektive ergänzt, das dient als Korrektiv der Lehrkräfteperspektive. Es wird auf die erste Sekundarstufe abgezielt, um auf formell oder informell erworbene digitale Kompetenzen aufbauen zu können. Die Erhebung und Auswertung erfolgt im Rahmen einer Grounded Theory Methodology[Ch14], um der Beleuchtung des weißen Flecks adäquat zu begegnen.

Zwei augenfällige Anforderungsideen, welche sich während des Initial Coding zeigen, sind adaptive Hilfestellungen für kooperative Lernformen und eines inklusiven Informationspools zur Begleitung der Schüler\*innen durch kooperative Lernformen. Sie dienen einerseits als gute Ausgangspunkte zur Theoriebildung. Andererseits eignen sie sich als erste Zielformulierungen, welche im weitere Participatory Design-Prozess zu konkreten Features ausgeschärft und als Prototyp implementiert werden können. Bemerkenswert ist die Erfahrung beider Lehrkräfte, dass Inklusion in der Schulpraxis durch Individualisierung (methodisch und Lernmaterial) als Spannung gegenüber kooperativen Lernformen wahrgenommen wird.

### 3 Ausblick

Im Fortgang des Forschungsvorhabens werden diese Ideen partizipativ durch Workshops ausgearbeitet. Anschließend werden sie als Prototyp implementiert und validiert<sup>3</sup>. Es werden dabei auch nach weiteren Bedingungen und Anforderungen für die „Individualisierung in der Gruppe“ gesucht.



### Literaturverzeichnis

- [Ch14] Charmaz, Kathy: Constructing grounded theory. 2nd. Aufl. Los Angeles : SAGE Publications Inc., 2014.
- [NKN19] Naoufal, Ala ; Konert, Johannes ; Naoufal, Hicham: EDUSPACE - Social Media Lernplattform mit Schwerpunkt Inklusion. In: Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft für Informatik (GI). Bd. P-297 : Gesellschaft für Informatik (GI), S. 293–294 , 2019.
- [PP19] Patzer, Yasmin ; Pinkwart, Niels: Inklusives Lernen mit barrierefreien Videos: Ein Fallbeispiel. In: Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft für Informatik (GI). Bd. P-297 : Gesellschaft für Informatik (GI), S. 311–312, 2019.

---

<sup>3</sup> Ein Beitrag zum Workshop “Digitally supported inclusive practices in education and training” ist geplant.

## Digitales Lesetheater zur Förderung der Leseflüssigkeit


Sara Durski <sup>1</sup>, Ute Massler<sup>2</sup> und Wolfgang Müller <sup>3</sup>

**Abstract:** In diesem Beitrag wird ein Konzept für ein digitales Leseflüssigkeitstraining vorgestellt. Mit diesem Programm haben GrundschülerInnen auch außerhalb des Schulunterrichts die Möglichkeit, das laute Lesen zu üben. Das Konzept basiert auf der Methode des Mehrsprachigen Lesetheaters, einer Lautlesemethode zur Förderung der Leseflüssigkeit.


**Keywords:** Lesetheater, Leseflüssigkeit, digitale Lernumgebung, e-Learning, spielbasiertes Lernen

Von 2001 bis 2016 stieg der Anteil der leseschwachen Viertklässler in Deutschland von 16,9% auf 18,9%. Hinzu kommt die wachsende Heterogenität der Schülerschaft [Bu17]. Eine zentrale Komponente einer guten Lesekompetenz ist die Leseflüssigkeit [Ga09]. Durch Verfahren des wiederholenden Lautlesens kann die Lesegenauigkeit, -flüssigkeit und -verständnis verbessert werden [Ni00]. In diesem Projekt wird ein digitales Lernmedium für die Methode des Mehrsprachigen Lesetheaters (MELT), welche auf dem Verfahren des wiederholten Lautlesens basiert, entwickelt und evaluiert. Im analog realisierten Lesetheater üben SchülerInnen kooperativ das Vorlesen von dialogischen Texten, indem sie in Sprecher- und Erzählerrollen aufgeteilt sind. Zum Abschluss werden die geübten Texte in gestaltender Form im Plenum aufgeführt [Ni06]. MELT ist eine Erweiterung des Lesetheaters um einen zwei- oder mehrsprachigen Ansatz, der von Lernenden und Lehrenden als sehr motivierend und lehrreich eingeschätzt wird [Ku17]. Um die Intensität des Übens der Leseflüssigkeit weiter zu steigern, bedarf es Selbstlernphasen. Die meisten in MELT genutzten Lautlesetechniken setzen jedoch einen Lesepartner voraus und dieser steht den Lernenden zuhause nicht immer zur Verfügung. Durch den Einsatz von geeigneten medialen Angeboten kann dies geleistet werden. Die Zielgruppe dieses Projekts sind leseschwache SchülerInnen der 3. und 4. Klassenstufe. Durch das digitale Lesetheater können SchülerInnen zuhause selbstgesteuert das Lautlesen mit virtuellen Lesepartnern üben. Zudem kann die Funktion des kooperativen Lesens mit andern SchülerInnen freigeschaltet werden. Das digitale Lesetheater orientiert sich an dem achtstufigen Lehr-/Lernprozess von MELT [Ku17]: (1) Einführung in das mehrsprachige Lesetheaterstück: Im Programm übernimmt ein Avatar die inhaltliche, sprachliche und technische Einführung. Zudem führt er den Lernenden durch das Programm. (2) Mehrsprachiges Vorlesen eines Einführungstextes zum Lesetheaterstück:

---

<sup>1</sup> PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, durski@md-phw.de,   
<https://orcid.org/0000-0003-0744-7983>

<sup>2</sup> PH Weingarten, Fach Englisch, Kirchplatz 2, 88250 Weingarten, massler@ph-weingarten.de

<sup>3</sup> PH Weingarten, Mediendidaktik, Kirchplatz 2, 88250 Weingarten, mueller@md-phw.de,   
<https://orcid.org/0000-0001-6474-3733>

Eine digitale Sprachausgabe gibt den Lesetext wieder, übernimmt dabei die Rolle des Sprach- und Lesemodells, wodurch der Lernende auch Charaktere, Konflikte und Setting des Lesetheaterstücks kennenlernt. (3) Lesen der Lesetheaterszene und Rollenverteilung: Der Lernende liest zunächst den dialogischen Lesetheatertext und wählt dann eine Leserolle aus. (4) Erarbeiten einer Inhaltsangabe der Lesetheaterszene: Um das Textverständnis zu überprüfen und ggf. weitere Maßnahmen zu ergreifen, dieses zu sichern, muss der Lernende bspw. verschiedene Ausschnitte der Szene inhaltlich ordnen. (5) Vorstellen der Szenen und Figurenrollen: Der Lernende kann das Aussehen seiner digitalen Rollenfigur verändern und spielerisch die Lernumgebung individuell gestalten, indem er Requisiten und Bühnenbilder für die Theateraufführung auswählt. (6) Üben und Feedback: Zum Üben der dialogischen Texte stehen dem Lernenden digitale Lesepartner, die durch verschiedene Tiere dargestellt werden, zur Verfügung. Diese Lesepartner übernehmen das Lesen von sämtlichen fehlenden Sprecher- und Erzählerrollen. Ergänzend dazu gibt das Programm dem Lernenden auf seine laut vorgelesene Rolle Feedback. Nach dem Üben in Selbstlernphasen kann der Lernende online gehen und mit anderen Lernenden live das Lesetheater lesen. Hierfür kann die Lehrperson u.a. die Gruppeneinteilung verwalten. (7) Generalprobe und Feedback: Hierdurch können die Lernenden in einem virtuellen kooperativen Setting die abschließende Aufführung ihrer Szene üben und sich gegenseitig Rückmeldung geben. (8) Aufführung: In dem Programm kann der Lernende seine Lesepformance aufnehmen. Nicht nur der Lernende selbst kann sich seine Aufnahmen anhören, sondern auch die Lehrperson, um ihren SchülerInnen Feedback zu geben. Im nächsten Schritt wird ein Prototyp des digitalen Lesetheaters mit GrundschülerInnen getestet. Hierbei wird durch Usability-Tests und ethnographische Interviews die Benutzerfreundlichkeit und Akzeptanz untersucht.

## Literaturverzeichnis

- [Bu17] Bundesministerium für Bildung und Forschung: Stabile Ergebnisse bei zunehmenden Herausforderungen – Lesen muss gestärkt werden, <https://www.bmbf.de/de/stabile-ergebnisse-bei-zunehmenden-herausforderungen---lesen-muss-gestaerkt-werden-5232.html>, Stand: 14.03.2020.
- [Ga09] Grabe, W.: Reading in a second language. Moving from Theory to Practice. Cambridge University Press, New York, 2009.
- [Ku17] Kutzelmann, S.; Massler, U.; Peter, K.; Götz, K.; Ilg, A. (Hrsg.): Mehrsprachiges Lesetheater. Handbuch zu Theorie und Praxis. Budrich, Opladen, 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.



## Teaching Machine Learning and Data Literacy to Students of Logistics using Jupyter Notebooks

Marvin Kastner<sup>1</sup>, Janna Franzkeit<sup>2</sup> and Anna Laine<sup>3</sup>

**Abstract:** Teaching machine learning in fields outside of computer sciences can be challenging when the students do not have a solid code knowledge. In this work, the requirements for teaching data literacy and code literacy to students of logistics are explored. Specifically, the use of Jupyter Notebooks in a machine learning course for students in logistics is evaluated, using “Teaching and Learning with Jupyter” written by Barba et al. in 2019 that lists several teaching patterns for Jupyter Notebooks.

**Keywords:** Jupyter Notebooks, Code Literacy, Data Literacy, Machine Learning, Data Science, Logistics, Supply Chain

### 1 Introduction

In the course “Machine Learning in Logistics”, students from logistics Master’s programs learn about data science and machine learning with a focus on their future workspace. A common definition of logistics is the accumulation of seven R (7-Rs): How to transport the right good in the right quantity and in the right quality at the right time to the right place in the right condition while this service can be offered at the right price [Co09]. During the lecture, concepts are taught while in the exercises practical skills are in focus, such as how to read in, explore, analyze, and visualize temporal and spatiotemporal data sets as well as large image data sets. Here, Jupyter Notebooks with Python code are used. Students enter the course with a very heterogeneous knowledge about programming partly because coding abilities are not a prerequisite for the course. Therefore, creating exercises that are appealing for both beginners and skilled programmers was a challenge.

A data scientist in logistics is expected to be both literate in code and data. Code literacy describes the ability to understand the underlying concepts of technology that surrounds us. In the scope of the presented course, no prerequisite in coding is required, for code is mainly seen as a tool to teach data literacy. Therefore, the students are mainly taught to read and modify existing code. Data literacy, on the other hand, represents the ability to understand and make decisions from data in their context. In the course, three main axes of this topic are explored with the use of coding exercises: collection of data from data

---

<sup>1</sup> Hamburg University of Technology, Institute of Maritime Logistics, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, marvin.kastner@tuhh.de

<sup>2</sup> Hamburg University of Technology, Institute of Maritime Logistics, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany, janna.franzkeit@tuhh.de

<sup>3</sup> Hamburg University of Technology, Institute for Software Systems, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany, anna.laine@tuhh.de

sources, performance of basic statistical analyses including problem solving with machine learning, and creation of visuals for dataset characteristics and experiment results.

Over the last two years, students had been continuously invited to provide feedback on the course, both orally and in anonymous, written form. Since it is a highly specialized course in the Master's programs, the number of participants at the end of the semester was not high enough to obtain statistically significant results from questionnaires. Nevertheless, the open questions lead to the conclusion that the general concept was addressing the expectations of the students. However, they did not comment on the style of the questions. In the exercise sheets, whenever students were asked to write code, interpret statistics, or describe graphs, this was done with Jupyter Notebooks. The exercise sheets were created without common guidelines by several research assistants. They displayed large individual differences in types and frequencies of specific patterns. This led to the idea of a detailed evaluation. Therefore, an overview that lists 22 different types of pedagogical patterns that can be used for teaching with Jupyter Notebooks [Ba19] was found. The patterns were analyzed regarding their ability to teach data literacy and 14 patterns were identified as useful in that regard. It was evaluated, that seven of them were used already, at least in parts, in the scope of the exercise. Furthermore, six patterns were identified, that can enrich the exercises in future and are thus planned to be applied newly or in a larger scope. To gain better insights, as well as evaluate the impact of the new patterns, more course iterations are required.

Jupyter Notebook for teaching code and data literacy to logistics students proved to be successful in that it allows a comfortable environment for the students to try out coding without the constraints of having to write a piece of code from the beginning to the end. In addition, visualizations and error messages are displayed close to the code that evoked it. This supports better the use of code as a tool for focusing on data literacy. This use of Jupyter Notebook allows beginners to stay in a safe environment when coding, without restricting experts from writing complex code.

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project "Maschinelles Lernen in Theorie und Praxis" was funded by the German Federal Ministry of Education and Research [grant number 01IS17077A].

## Bibliography

- [Ba19] Barba, L. A. et al.: Teaching and Learning with Jupyter. <https://jupyter4edu.github.io/jupyter-edu-book/>, accessed 21 Feb 2020.
- [Co09] Coyle, J. J.: Supply chain management. A logistics perspective. South-Western, Mason, Ohio, 2009.

## Bearbeiten Informatiklehrkräfte außerhalb der Dienstzeit ein Blended-Learning-Modul für eine Lehrerfortbildung?

Kensuke Akao<sup>1</sup>

**Abstract:** Viele Informatiklehrkräfte geben mangelnde Zeit und lange Fahrzeiten als Gründe an, um an Fortbildungen nicht teilzunehmen. Deshalb besteht in diesem Bereich die Überlegung, Blended-Learning einzusetzen. Die meisten Informatiklehrkräfte in Nordrhein-Westfalen wünschen sich jedoch schulexterne Fortbildungen innerhalb der Dienstzeit. Dieser Beitrag wird empirische Evaluationsergebnisse zum Einsatz von Blended-Learning in einer Informatiklehrerfortbildung vorstellen, ob die Teilnehmende ihre Freizeit benutzen, um die Online-Phase zu bearbeiten.

**Keywords:** Informatik, Lehrer, Fortbildung, Blended-Learning, Medien, Lernmaterial, E-Book

### 1 Flexibilität durch Online-Lernen vs. „Freizeit ist heilig“

In Nordrhein-Westfalen (NRW) sind die Lehrkräfte gesetzlich dazu verpflichtet, sich in ihrer unterrichtsfreien Zeit regelmäßig fortzubilden (vgl. § 57 Abs. 3 SchulG NRW). Allerdings sehen viele Lehrkräfte Hürden bezüglich Zeit und Ort, um an Fortbildungen teilzunehmen. Auf der einen Seite lässt sich hervorheben, dass der verstärkte Einsatz von Lehrerfortbildungsangeboten via E-Learning die Unabhängigkeit von Zeitpunkt und Ort ermöglicht [Gr15]. Auf der anderen Seite wissen wir aufgrund unserer vorangegangenen Forschungsergebnisse jedoch, dass sich die meisten Informatiklehrkräfte in NRW eine Fortbildung innerhalb der Dienstzeit wünschen. Eine interessante Begründung von Lehrkräften ist, dass Freizeit heilig ist. D. h., die Flexibilität von E-Learning oder Blended-Learning in der Online-Phase als 24 Std.-Lernumgebung scheint durch die Bereitschaft deutscher Lehrkräfte begrenzt. Dieses Ergebnis warf die Frage auf, wann Informatiklehrkräfte sich hauptsächlich mit dem Online-Lernen der Fortbildung beschäftigen. Unser Projekt namens FIT-2 bietet Informatiklehrerfortbildungsmodul unter dem Einsatz von Blended-Learning basierend auf den Methoden *Mikrolernen* [Hu10] an. Dabei wenden wir unser Konzept der Online-Phase „Video-Text-Hybridlehrmaterial“ in dem Blended-Learning-Modul an. Entwickeltes Material wurde es am 25. März 2019 in einem Lehrerfortbildungsworkshop während der Fachtagung NRW-Informatiktag 2019<sup>2</sup> verwendet, damit das Potenzial des Hybridmaterials, welches unser Vorschlag für eine bessere Akzeptanz der Online-Module in Blended-Learning ist, mithilfe der Umfrage evaluiert wurden. Die Anzahl der angemeldeten Teilnehmenden war 48 Personen, 35 Personen nahmen tatsächlich an der Fachtagung teil [Ak20].

---

<sup>1</sup> Westfälische Wilhelms-Universität (WWU) Münster, Didaktik der Informatik (DDI), Corrensstraße 80,  
48149 Münster, kensuke.akao@uni-muenster.de

<sup>2</sup> <https://informatiktag-nrw.de> (zuletzt geprüft am: 10.2.2020)

## 2 Ergebnis

Die Teilnehmenden hatten 20 Tage Zeit, um die Online-Phase zu bearbeiten. Aus der Mehrfachauswahlfrage „Wann haben Sie sich hauptsächlich mit der Online-Phase beschäftigt?“ wurde deutlich, dass nur eine Person der 18 Teilnehmenden an Werktagen in der Dienstzeit das Modul der Online-Phase bearbeiten konnte. Allerdings hatte sie nicht genug unterrichtsfreie Zeit im Dienst, um alle Aufgaben fertig zu stellen, deshalb glich sie am Wochenende ihren Zeitmangel für das Lernen aus. 11 von den verbleibenden 17 Teilnehmenden bearbeiteten die Online-Module am Wochenende und weitere fünf Teilnehmende an Werktagen außerhalb der Dienstzeit und zuletzt eine Person nicht nur an Werktagen außerhalb der Dienstzeit, sondern auch am Wochenende. Obwohl bei der Bearbeitung die Freizeit benutzt wurde, bearbeiteten 16 der 18 Befragten den ersten und zweiten Teil vollständig, die übrigen zwei Befragten bearbeiteten nur den ersten Teil. Außerdem bearbeiteten 11 der 18 Teilnehmende eine freiwillige Übung (eine Person bearbeitete nur den ersten Teil). Über den Zeitaufwand für die Online-Phase (15 min pro Teil) zur kompakten Fortbildung bei geringer zeitlicher Verfügbarkeit in der unterrichtsfreien Zeit sagten sieben der 18 Teilnehmende, sie sei „Sehr hilfreich (5)“, weitere neun Teilnehmende „hilfreich (4)“ und letzte zwei 2 Teilnehmende „Mittelmäßig (3)“. Niemand wählte „Nicht hilfreich (2)“ oder „Gar nicht hilfreich (1)“. Schließlich fragten wir danach, ob Blended-Learning oder reines E-Learning (ohne Präsenz) im Bereich der Informatiklehrerfortbildungen wirkungsvoll eingesetzt werden kann (n=10). Acht Teilnehmende stimmen vollkommen oder stimmen zu, dass Blended-Learning wirkungsvoll eingesetzt werden kann. Dabei bekamen wir die Rückmeldungen, dass unsere Online-Phase mit dem Video-Text-Hybridmaterial sich auf die Teilnahme positiv auswirkt. Aber dem reinen E-Learning stimmen nur drei Teilnehmende zu, während fünf Teilnehmende dagegen sind. Dabei sind sechs Teilnehmende der Meinung, dass der Austausch oder die aktive Diskussion fehlen, wenn eine Lehrerfortbildung nur als reine E-Learning angeboten würde. Deshalb kann das von uns entwickelte Konzept der Blended-Learning-Szenarios eine Möglichkeit sein, um die Akzeptanz von Lehrkräften zu gewinnen.

## Literaturverzeichnis

- [Ak20] Akao, K.: Ergebnis einer Blended-Learning-Informatiklehrerfortbildung mit dem Einsatz von Video-Text-Hybridlehrmaterial. In (Schuhen, M. & Froitzheim, M., Hrsg.): Das Elektronische Schulbuch 2020. Fachdidaktische Anforderungen und Ideen treffen auf Lösungsvorschläge der Informatik. LIT Verlag, Münster, 2020 (im Druck).
- [Gr15] Grabner, G.: Blended Learning - blendende Aussichten oder nur Blendwerk?: Entwicklungsperspektiven in der Lehrer/innenfortbildung. S. 18, 2015.
- [Hu10] Hug, T.: Mikrolernen – konzeptionelle Überlegungen und Anwendungsbeispiele. In: Jahrbuch Medienpädagogik 8. VS Verlag für Sozialwissenschaften, Wiesbaden, S. 221-238, 2010.

## Learning Analytics und Diskriminierung

Marius Wehner<sup>1</sup>, Alina Köchling <sup>2</sup>

**Abstract:** Learning Analytics (LA) wird immer häufiger als neue Beratungsquelle für Lehrende eingesetzt. Bildungseinrichtungen setzen LA in ihrem Benotungsprozess ein, um die Effizienz sowie die Objektivität zu steigern. LA kann jedoch auch zu einer ungerechten Behandlung bestimmter Personengruppen führen und somit zu impliziter Diskriminierung. Ziel der geplanten Studie ist es das Diskriminierungspotenzial mit einer experimentellen Conjointanalyse aufzuzeigen, um somit Bewusstsein für die negativen Folgen von LA zu schaffen.

**Keywords:** Learning Analytics, Discrimination, Conjoint Analysis, E-Learning, Fairness

### 1 Einleitung

Ziel von LA es, durch dynamisches, interaktives Lernen den Lernerfolg positiv zu beeinflussen. Neben diversen Vorteilen birgt der Einsatz von LA auch einige Risiken, die bei der Implikation in den schulischen Alltag berücksichtigt werden sollten. Kernpunkt sind hierbei diverse Diskriminierungspotentiale im Kontext der algorithmischen Datenauswertung. Algorithmen treffen ihre Entscheidungen auf Grundlage von Trainingsdaten. Enthalten diese Datensätze verzerrte Informationen, so werden auch die Empfehlungen der Lernplattform verzerrt sein (vgl. [DF18]). Derartige Verzerrungen können sich auf demographische Aspekte der Lernenden (Geschlecht, Herkunft) und auf lernbezogene Aspekte (Lerntyp, Nutzung von digitalen Plattformen) beziehen (vgl. [Sc16], [SH15]). Bisher analysieren nur wenige Studien das Risiko der Diskriminierung bei der Anwendung von LA. Die Idee dieser Forschungsarbeit besteht daher darin, den Einsatz von digitalen Lernplattformen und LA auf mögliche Diskriminierungspotentiale bei der Leistungsbeurteilung der Lernenden hin zu untersuchen.

### 2 Studie und Methode

Um die mögliche Diskriminierung von Lernenden durch Lehrende bei der Leistungsbeurteilung, die durch verzerrte Informationen der digitalen Lernplattform beeinflusst wird, zu untersuchen, wird eine adaptive choice-based Conjoint-Analyse (ACBC) durchgeführt. In dieser Studie wird das erprobte, computergestützte Verfahren zur Analyse von komplexen Entscheidungsprozessen dazu verwendet, die relevanten

---

<sup>1</sup> Heinrich-Heine-Universität Düsseldorf, Juniorprofessur für BWL, insb. Unternehmensführung,  
Universitätsstraße 1, 40225 Düsseldorf, marius.wehner@hhu.de

<sup>2</sup> Heinrich-Heine-Universität Düsseldorf, Juniorprofessur für BWL, insb. Unternehmensführung,  
Universitätsstraße 1, 40225 Düsseldorf, alina.koechling@hhu.de, <https://orcid.org/0000-0001-7039-9852>

Entscheidungsfaktoren zu ermitteln, auf Grundlage derer Lehrende eine Leistungsbeurteilung von Lernenden vornehmen. Als Probanden werden Lehrende aus deutschen weiterführenden Schulen befragt. Der verwendete Fragebogen beinhaltet ein fiktives Szenario der Leistungsbeurteilung. Darin enthalten sind 16 Auswahlaufgaben, bei denen jeweils zwei fiktive Lernende präsentiert werden. Der Lehrende erhält nun die Aufgabe, denjenigen Lernenden auszuwählen, den er als leistungsstärker beurteilt. Die Beschreibung der Lernenden erfolgt anhand einheitlicher Attribute und Attributausprägungen (Levels).

Attribut	Ausprägung
Foto	Bild (durch KI generiert)
Name	Alexander, Mohammed, Marie, Sophie, Layla, Elif zur Analyse von komplexen Entscheidungsprozessen
Lerntyp	gar nicht, immer vor Klausur, permanent (Lerntyp als Aktivitätskurve dargestellt)
Onlinetests	3/19, 7/19, 12/19
Unterstützung	kaum, mäßig, viel
Plattformempfehlung	gefährdet, positiv ausgezeichnet

Tab. 1: Attribute und Attributausprägungen der Conjoint-Analyse

Die beschriebene ACBC-Analyse ermittelt nach Befragung aller Probanden relative Attributwichtigkeiten. Diese beschreiben als Prozentwert, inwieweit das jeweilige Attribut zur Entscheidungsfindung beigetragen hat. So kann zum Beispiel bestimmt werden, wie stark Lehrende den Lerntyp oder den Namen des Lernenden mit in ihre Entscheidung aufgenommen haben. Ferner lassen sich auch individuelle Nutzenwerte der Probanden messen. Diese geben Auskunft darüber, welche spezifischen Level bevorzugt wurden. Auf aggregierter Ebene ist es somit möglich, verschiedene Diskriminierungspotentiale zu identifizieren. Fraglich ist beispielsweise, ob Lernende, die zwar viel analog lernen, aber nur selten die Lernplattform verwenden (Lerntyp "gar nicht"), bei der Leistungsbeurteilung gegenüber anderen Lernenden benachteiligt werden oder ob die (möglicherweise fehlerhaften) Empfehlungen der Plattform die tatsächliche Benotung beeinflusst.

## Literaturverzeichnis

- [DF18] Dressel, J., Farid, H.: The accuracy, fairness, and limits of predicting recidivism. *Science advances* 4/1, eaao5580, 2018.
- [Sc16] Scholes, V.: The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development* 64/5, 939-955, 2016.
- [SH15] Shahiri, A. M., Husain, W.: A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, 72, 414-422, 2015.

## Virtual Reality-based Digital Reusable Learning Objects in Healthcare Training (ViRDIPA): 360°-Video Trainings

Leonard Meyer<sup>1</sup>, Thies Pfeiffer<sup>2</sup>

**Abstract:** Training and education in Virtual Reality has become a widely discussed field in research and teaching. One decisive aspect is the efficient creation of content. Especially creating detailed virtual 3D environments and integrating specific training procedures is usually time consuming and/or expensive. Furthermore, these scenarios often cannot be created by the trainers themselves. The goal of the project is therefore to enable instructors and teachers to quickly create their own trainings using pre-created content (Digital Reusable Learning Objects) and 360°-images for easy virtual environment creation. Based on a developed and evaluated concept, an authoring tool is to be created to enable this generation of educational content in the field of healthcare. Important parts of the development are enabling of multi-user interaction and integration into existing learning management systems.

**Keywords:** Virtual Reality Training; Digital Reusable Learning Objects; 360°-VR; Authoring Tool

### 1 Motivation and Project Embedding

On the largest scale, the project 'Virtual Reality-based Digital Reusable Learning Objects in Healthcare Training' aims to develop, test and evaluate a blended learning training concept for the use of virtual reality technology in healthcare education. Accordingly, traditional class-based teaching methods should be combined with online education materials and interaction possibilities. Virtual fully immersive first-person perspective trainings promise to significantly increase the learning success in many situations due to the high level of reality reference that is possible. Nevertheless, there are still various uncertainties regarding efficiency in relation to the respective area of application and problems with the integration of such systems.

The main approach in this project is the use of digital reusable learning objects (DRLOs) to enable an efficient creation of educational VR content and thus provide a simple-to-use authoring tool for teachers. The VR learning materials created should be available online for students and enable cooperative work, possibly guided by the lecturer. Therefore, integration into existing learning management systems (LMS) is an important aspect for universal applicability. To achieve the set goals, an interdisciplinary team consisting of the departments of healthcare, didactics, computer science and innovative media will work together over a period of three years.

---

<sup>1</sup> Hochschule Emden/Leer, Abteilung E+I, Constantiaplatz 4, 26723 Emden, leonard.meyer@hs-emden-leer.de,

<sup>2</sup> Hochschule Emden/Leer, Abteilung E+I, Constantiaplatz 4, 26723 Emden, thies.pfeiffer@hs-emden-leer.de

## 2 Project Approach to 360°-Video-based Trainings

Building up a fully virtual 3D environment is usually a time-consuming and expensive task, especially if the results are to be on a realistic level. In comparison, the recording of 360° videos and their integration into a VR scene takes much less time. In addition, the content creation part is then primarily conducted in the real environment, which is much more approachable for the target audience. Studies investigating learning outcomes support the assumption that 360°-videos are an effective extension of learning [TL19] [Ch19]. Therefore, the planned authoring tool should enable content creators to record their own 360°-videos of procedures and use them as a basis for VR learning materials. The possibility of integrating interactive digital elements within the 360° (VR) video (similar to Augmented Reality in a real context) allows a more complex storyline to be created [Rh17], for instance using multiple choice decisions influencing the training in progress. The pure reception of information is thus extended by an interactive component.

**Research questions addressed in the project are:** How are the generated healthcare contents accepted by the trainees? What is the level of fidelity when trainers use the authoring tool? Does the developed Concept improve learning success as well as the integration of 360°-VR in an educational context?

## 3 Outlook for the Poster

The poster will give an overview of the project goals, briefly reports on results from several assessments currently underway to collect requirements from the different stakeholders involved in the context of training (administration, teachers, students) and focuses on the presentation of the first concept for the authoring workflow.

The project 'ViRDIPA' is funded by the Bundesministerium für Bildung und Forschung.

## Bibliography

- [Ch19] Chang, Ching-Yi; Sung, Han-Yu; Guo, Jong-Long; Chang, Bieng-Yi; Kuo, Fan-Ray: Effects of spherical video-based virtual reality on nursing students' learning performance in childbirth education training. *Interactive Learning Environments*, pp. 1–17, 2019.
- [Rh17] Rhee, Taehyun; Petikam, Lohit; Allen, Benjamin; Chalmers, Andrew: Mr360: Mixed reality rendering for 360 panoramic videos. *IEEE transactions on visualization and computer graphics*, 23(4):1379–1388, 2017.
- [TL19] Taylor, Natasha; Layland, Adam: Comparison study of the use of 360-degree video and non-360-degree video simulation and cybersickness symptoms in undergraduate healthcare curricula. *BMJ Simulation and Technology Enhanced Learning*, 5(3):170–173, 2019.



## Your Lesson, Your Way, Your Success

### Kompetenzvermittlung im Inverted Classroom Mastery Modell

Annett Thüring<sup>1</sup> und Kathrin Jäger<sup>2</sup>

**Abstract:** Aus den Überlegungen heraus mit modernen zeitgemäßen Lehr-/Lernkonzepten dem universitären Anspruch auf Höchstqualifikation gerecht zu werden, sowie vor dem Hintergrund der steigenden Heterogenität der Studierenden bzgl. des individuellen Vorkenntnisstandes, wurde der Kurs Objektorientierte Programmierung (OOP) im Inverted Classroom Mastery Modell erstellt. Das Konzept berücksichtigt zusätzlich die Bedürfnisse und Wünsche der Studierenden nach individualisierten Studienwegen und flexiblen Studienmodellen. Vorgestellt wird damit eine lernerzentrierte Kursumsetzung, die didaktisch auf eine stufenweise Vermittlung von Fachkompetenz fokussiert, selbstreguliertes Lernen fördert und den Lernerfolg jeder Stufe ermöglicht, misst und transparent sowohl für den Lehrenden als auch den Studierenden sichtbar macht.

**Keywords:** Inverted Classroom Mastery Modell, Kompetenzmessung, Programmierung

Hohe Anforderungen des Studiums, fehlende fachliche Voraussetzungen aber auch seitens der Studierenden ein zu geringer Praxisbezug sind häufige Ursachen für einen in den letzten Jahren zunehmenden frühen Studienabbruch [He17]. Mit „Your Lesson, Your Way, Your Success“ fokussierte eine Umstrukturierung der Lehrveranstaltung OOP auf eine stärker lernerzentrierte stufenweise Wissensvermittlung mit formativen Assessments in Online-Phasen und auf einen hohen Praxisbezug in Präsenzübungen. Erste Praxiserfahrungen basierend auf dem dreiphasigen Inverted Classroom Mastery Modell (ICMM) [Ha15] für das Modul OOP können positiv diskutiert werden.

Das Konzept des ICMM zielt darauf ab, die Wissensvermittlung online vorzulagern und Assessments während dieses Lernprozesses unterstützend zu integrieren sowie die Präsenzzeit zur Übung und zum intensiven Diskurs zu nutzen. Die Rolle des Lehrenden verändert sich in der Struktur des ICMM zum Lernprozessbegleiter, orientiert auf das Lernen der Studierenden. Die Abbildung zeigt neben der Grundstruktur der Inhaltsvermittlung einer der 15 Kurseinheiten aus Distanz- und Präsenzphase, auch die Kommunikationselemente, die den Lernprozess unterstützen.

---

<sup>1</sup> Martin-Luther-Universität Halle-Wittenberg, Institut für Informatik, Von-Seckendorff-Platz 1, 06120 Halle (Saale), annett.thuering@informatik.uni-halle.de

<sup>2</sup> Martin-Luther-Universität Halle-Wittenberg, Zentrum für multimediales Lehren und Lernen (LLZ), Hoher Weg 8, 06120 Halle (Saale), kathrin.jaeger@llz.uni-halle.de



## **Demos**



## Entwicklung des „KI-Campus – Die Lernplattform für Künstliche Intelligenz“

Dana-Kristin Mah<sup>1</sup>, Cornelia Gamst<sup>1</sup>, Lavinia Ionica<sup>1</sup> und Christian Dufentester<sup>1</sup>

**Abstract:** Vorgestellt wird der prototypische Aufbau einer auf das Thema Künstliche Intelligenz spezialisierten digitalen Lernplattform. Lernende und Lernprozesse stehen im Mittelpunkt der Plattform – dazu wird der Prototyp mit zukunftsfähigen Lernformaten und KI-Anwendungen nutzerzentriert entwickelt.

**Keywords:** Digitale Lernplattform, Künstliche Intelligenz, innovative, digital gestützte Lehr- und Lernformen, Open Educational Resources, Learning Analytics, Personalisierung

### 1 Digitale Lernplattformen und Künstliche Intelligenz

Digitale Lernplattformen bieten Lernenden die Möglichkeit flexibel zeit- und ortsunabhängig zu lernen. Neben relevanten und qualitativ hochwertigen Lerninhalten sind didaktisch innovative und zukunftsfähige Lehr- und Lernformate von zentraler Bedeutung, um Lernenden mit ihren individuellen Bedarfen und Lernerfahrungen gerecht zu werden. Beispiele für digitale Lernplattformen auf internationaler Ebene sind etwa Coursera und edX, in Europa z. B. FutureLearn und FUN und in Deutschland OpenHPI, HOUU und oncampus. Bildungsangebote zum Thema Künstliche Intelligenz (KI) sind derzeit vorwiegend auf internationalen Plattformen verfügbar, in Europa ist insbesondere der finnische Online-Kurs Elements of AI verbreitet (Mah & Büching, 2019). In Deutschland sind vergleichbare Bildungsangebote auf einer digitalen Lernplattform speziell für Themen der KI bisher noch nicht vorhanden (Schmid et al., 2018). Künstliche Intelligenz verändert jedoch bereits heute unsere Lebenswelt und wird zukünftig noch wichtiger werden. In der „Strategie Künstliche Intelligenz der Bundesregierung“ wird der Aufbau einer eigenen Lernplattform angeführt, die sowohl KI-Kompetenzen vermittelt, als auch KI-Elemente nutzt. (Die Bundesregierung, 2018). Das Pilotprojekt „KI-Campus – Die Lernplattform für Künstliche Intelligenz“ wird durch den Stifterverband, das Deutsche Forschungszentrum für Künstliche Intelligenz (DFKI), das Hasso-Plattner-Institut (HPI), NEOCOSMO und das mmb Institut umgesetzt.

---

<sup>1</sup> Stifterverband für die Deutsche Wissenschaft, Projekt KI-Campus, Tempelhofer Ufer 11, 10963 Berlin, {dana-kristin.mah, cornelia.gamst, lavinia.ionica, christian.dufentester}@stifterverband.de

## 2 KI-Campus – Die Lernplattform für Künstliche Intelligenz

Ziel des Vorhabens ist es, einen mündigen und kompetenten Umgang mit KI zu stärken und mittelfristig auch einem akuten Fachkräftemangel zu begegnen. Der KI-Campus adressiert dabei Studierende, Berufstätige und andere lebenslang Lernende, die sich für KI interessieren und auf diesem Gebiet weiterbilden möchten. Er richtet sich außerdem an Hochschulen, Forschungseinrichtungen und Unternehmen, die sich mit der Produktion eigener Lernangebote beteiligen oder Inhalte des KI-Campus in ihre Lehre integrieren möchten.

Leitprinzipien für die prototypische Entwicklung des KI-Campus umfassen unter anderem eine hohe Übersichtlichkeit, Personalisierbarkeit und Adaptivität der Plattformangebote sowie zukunftsfähige und innovative didaktische Konzepte, in deren Mittelpunkt die Lernprozesse der Lernenden stehen. Ziel der Lernplattform ist es, anhand einer thematisch breit angelegten Sammlung von Lernangeboten zu KI, KI-Kompetenzen zu vermitteln und mehr Menschen für das Thema zu begeistern. Über Zugänge durch Themenportale, Selbsttests oder filterbare Suche können die adressierten Personen entscheiden, ob sie grundlegende Konzepte und Anwendungsgebiete von KI kennenlernen, fach- und berufsspezifische Kompetenzen ausbauen (z.B. KI in der öffentlichen Verwaltung), oder in die konkrete Anwendung und Mitgestaltung einsteigen möchten (z.B. durch das Programmieren von Neuronalen Netzen). Die Lernangebote reichen dabei von offenen Selbstlernkursen, über zertifizierte Online- und Blended-Learning Formate, bis hin zu curricular verankerten bzw. anrechenbaren und mit ECTS vergüteten Studienmodulen. Um Lernenden das bestmögliche Angebot zur Verfügung zu stellen, werden zum einen eigene Lerninhalte im Rahmen von Wettbewerben und Kooperationen speziell für den KI-Campus entwickelt und zum anderen bestehende (offene) Lerninhalte integriert. Hochschulen und Unternehmen profitieren von der gemeinsamen Infrastruktur und der verbesserten Sichtbarkeit ihrer Lernformate. Für die Personalisierung werden innovative KI-basierte Anwendungen und Werkzeuge entwickelt und integriert. Adaptives Lernen, Lehren und Prüfen soll ermöglicht werden. Learning Analytics werden datenschutzkonform und mit dem Ziel der individuellen Unterstützung von Lernenden eingesetzt. Die Beta-Version der digitalen Lernplattform KI-Campus wird voraussichtlich Mitte 2020 als Prototyp verfügbar sein.

### Literaturverzeichnis

- [Di18] Die Bundesregierung: Strategie Künstliche Intelligenz der Bundesregierung. Berlin, 2018.
- [MB19] Mah, S.-K.; Büching, C.: Künstliche Intelligenz in Studium und Lehre. Überblickstudie zu Professuren und Studiengängen der Künstlichen Intelligenz in Deutschland. VDI/VDE Innovation + Technik GmbH, 2019.
- [Sc18] Schmid, U. et al.: Machbarkeitsstudie für eine (inter-)nationale Plattform für die Hochschullehre. Arbeitspapier Nr. 33. Hochschulforum Digitalisierung beim Stifterverband für die Deutsche Wissenschaft e.V., Berlin, 2018.

## Zirkus Empathico 2.0

### A serious mobile game for empathy enhancement in children with Autism

Ahmed Hassan<sup>1</sup> and Niels Pinkwart<sup>1</sup>

**Abstract:** The aim of the mobile app ‘Zirkus Empathico 2.0’ is to improve emotional empathy and social competencies in Pre-school children with autism spectrum disorder (ASD). The holistic idea of grounding is based on genuine results of empathy research. An examination of the app exposed its good usability and comprehensibility. The application ‘Zirkus Empathico 2.0’ is scheduled to be thoroughly investigated in a longitudinal clinical dissection in children aged five to ten.

**Keywords:** Serious games, Autism spectrum disorder, Social behavior in Autism, Emotional intelligence

## 1 Introduction and Novelty

Empathy is the ability to recognize and understand the feelings of other people [Dz08]. Several studies show that individuals suffering from autism have cognitive-reduced empathy, that’s why they have difficulty in understanding the emotions or gestures of the other people [Bo13]. Serious games (SG) approach to enhancing empathy through recognizing and producing precise facial emotions has proven to be very useful over the past two decades [Ha10]. Games that have a specific learning goal apart of entertainment are called SG. This article presents a SG that has been developed to enhance empathy in children with autism. Both the usability and design of SG are equally important and there is a need for a better understanding between developmental psychology and computer science. In ‘Zirkus Empathico 2.0’, we have considered these issues and added some useful features in comparison to its previous version ‘Zirkus Empathico’. A user profile is created to keep track of user performances, and the on-screen time. The game adapts to the individual user’s playing performance. An online and offline multiplayer module has been added to boost the social interaction and collaborative skills of the individuals. This SG is the first multi-player game for children with ASD of its sort.

## 2 Game Concept and Architecture

The mobile application ‘Zirkus Empathico 2.0’ consists of four distinct modules based on social cognition. Tasks are laid out to provoke individuals to determine their feelings or

---

<sup>1</sup> Department of Informatics, Rudower Chaussee 25, Room 3.405, 12489, Humboldt University Berlin, Germany, {hassanah, niels.pinkwart}@informatik.hu-berlin.de

emotions (module 1). The identification of other people's emotions from videotaped facial expressions (module 2). An individual's emotions in response to the other person's emotions must be defined for building up emotional empathy (module 3). To improve collaborative skills, there is a multiplayer mode where children play with their peers online or offline (module 4). Fig 1 explains the game concept. Game is designed in Unity 3D with modular approach. Cordova plugins were used for internal app logics. For the structure of application Model view controller (MVC) paradigm was used. There were total six view controller which were managed by tool grunt.

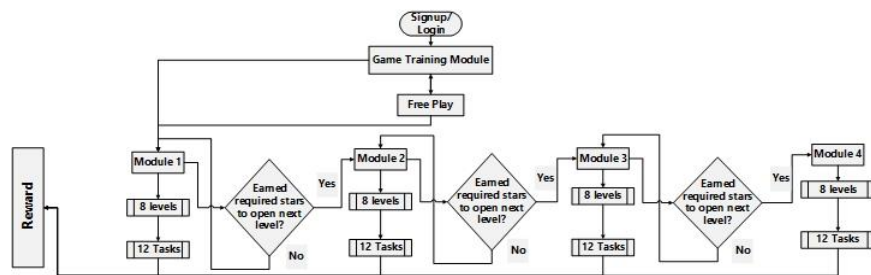


Fig. 1: Game concept and training module

The mobile application Zirkus Empathico 2.0 is created with multi-disciplinary research on developmental psychology and software engineering [BCH11]. This SG is the first multi-player game for children with ASD of its sort. After the scheduled clinical execution, we can investigate its viability in more detail.

## Literaturverzeichnis

- [Bo13] Bons, D. et al.: Motor, Emotional, and Cognitive Empathy in Children and Adolescents with Autism Spectrum Disorder and Conduct Disorder." *Journal of abnormal child psychology* 41.3, pp. 425-43, 2013
- [BCH11] Boyle, E.; Thomas M. C.; Thomas H.: The Role of Psychology in Understanding the Impact of Computer Games. *Entertainment Computing* 2.2, pp. 69-74, 2011.
- [Dz08] Dziobek, I. et al.: Dissociation of Cognitive and Emotional Empathy in Adults with Asperger Syndrome Using the Multifaceted Empathy Test (Met). *Journal of autism and developmental disorders* 38.3, pp. 464-73, 2008.
- [Ha10] Harms, M.B.; Alex M.; Gregory L. W.: Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies. *Neuropsychology review* 20.3, pp. 290-322, 2010.



## TABBI - A Tangible Interface for Educational Chatbots

Sebastian Wollny<sup>1</sup>, Jan Schneider<sup>1</sup>, Marc Rittberger<sup>1</sup> and Hendrik Drachsler<sup>1</sup>

**Abstract:** With the increasing popularity of Artificial Intelligence in educational research, chatbots seem to be a promising future technology that could help students with their daily problems. As part of Smart Learning Environments, they can be permanently accessible by natural language via a variety of interfaces. To investigate the property of "tangibility", we present TABBI, an interface for educational chatbots that allows students to interact with a physical robot. This robot interface will be used in future comparative studies to evaluate interaction possibilities with educational chatbots. In this paper, we describe the general idea and context of TABBI.

**Keywords:** Artificial Intelligence, Smart Learning Environments, Chatbot Interface, Tangibility, Robot, Assistant

### 1 Motivation

With possibilities of modern hardware, applications of Artificial Intelligence (AI) in education have become realistic in recent years. Chatbots in particular could help to improve the quality and availability of digital education. They enable natural language interactions by recognizing user intentions through an underlying neural network.

Chatbots can be integrated into a variety of applications and interfaced through text-stream inputs and outputs e.g. by instant messengers, websites, or smart speakers. In the field of educational technology research, interactions with chatbots have been researched for decades. However, only a few comparative studies of interfaces have been conducted in this regard so far. To enable future studies on the presence of chatbots, we have designed TABBI. It integrates into a widely used chatbot framework and can be utilized without changing the underlying logic. This should lead to more insights into robotic chatbots, smart speakers, and similar interfaces in the field of educational technologies.

### 2 Educational Setting

TABBI is a chatbot interface integrated into a robot, which can be placed in Smart Learning Environments to help students. According to the work of [Wo20] and [WS18], chatbots could lead to more optimized interactive Q&A systems, digital companions, or discussion partners. In this context, a number of pedagogical questions arise. The outstanding feature of TABBI, the physical embodiment, or "tangibility", can be used to

---

<sup>1</sup> 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

examine the significance of this feature in the educational context. Tangibility could lead to more natural or less distracting interactions on the one hand, but also to barriers, e.g. through a loss of privacy in public conversations. Experience in this respect could in turn be used to optimize chatbots and interfaces for their specific use in education and learning environments.

### 3 Technical Design

From a technical point of view, TABBI consists of a double robot 2, paired with a customer-specific connection to RASA, a chatbot framework commonly used in the educational sector. This connector consists of a sensor bridge for coupling the mobile robot device with the rest of the logic, a Text-To-Speech module (Mozilla TTS<sup>2</sup>), and an Automatic-Speech-Recognition module (Mozilla DeepSpeech<sup>3</sup>). A detailed elaboration of the interface design can be found in Fig. 1:

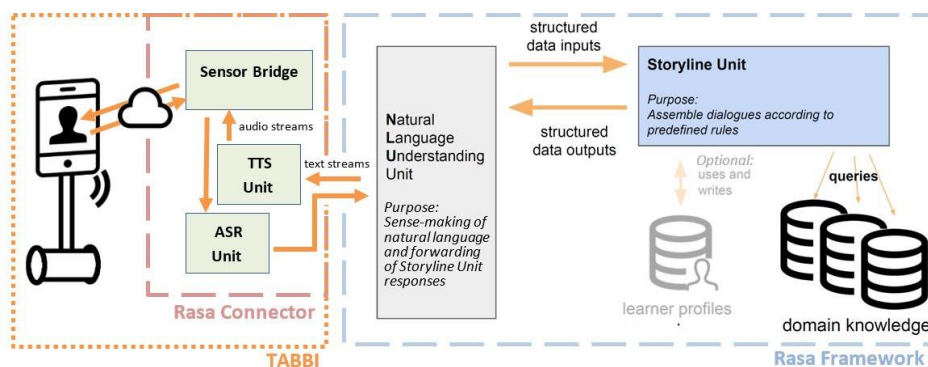


Fig. 1: Conceptual design of the robot chatbot interface TABBI

This structure makes it easy to exchange TABBI with other chatbot interfaces in order to compare human interactions based on the interface used.

### Bibliography

- [Wo20] Wollny, S.; Schneider, J.; Rittberger, M.; Drachsler, H.: Chatbots – An Opportunity for Individual Assistance in Education. Companion Proceedings of the 10<sup>th</sup> International Conference on Learning Analytics & Knowledge LAK20
- [WS18] Winkler, R.; Söllner, M.: Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. Academy of Management Annual Meeting (AOM) 2018

<sup>2</sup> Project website: <https://github.com/mozilla/TTS> – accessed on 2020-06-29, 8:36 pm

<sup>3</sup> Project website: <https://github.com/mozilla/DeepSpeech> – accessed on 2020-06-29, 8:35 pm

## LernBar 4.6: Barrierearme, für Learning Analytics optimierte Web Based Trainings

Sarah Voß-Nakkour<sup>1</sup>, Patrick Sacher<sup>1</sup>, Thorsten Gattinger<sup>1</sup> und David Weiß<sup>1</sup>

**Abstract:** Das Autor\*innensystem LernBar dient der Erstellung von interaktiven Lernmodulen mit methodisch-didaktischer als auch technischer Unterstützung. Es wird von **studiumdigitale**, der zentralen eLearning-Einrichtung der Goethe-Universität, entwickelt und besteht aus den LernBar-Komponenten: Studio (Kurserstellung als Desktopanwendung), Player (Wiedergabe der Lernmodule im Browser) und Portal (Webportal zur Distribution). Durch die Möglichkeit, didaktische Strukturen mit vorgegebener Gestaltung und Navigation einfach umzusetzen, hat sie sich an (Hoch-)Schulen, öffentlichen Trägern und Unternehmen zur Erstellung von Web Based Trainings etabliert. Die LernBar ist ebenso Grundlage für Forschungsprojekte (z.B. [WE18]) und individuelle Entwicklungen (z.B. im Bereich Learning Analytics, adaptives Testen, UX-Design). Im letzten Jahr gab es im agilen Softwareentwicklungsprozess zwei Schwerpunkte, die beforscht und umgesetzt wurden: Eine Erhöhung der Barrierefreiheit der Lernmodule und die Verbesserung der Erfassung sowie Auswertung von Nutzungsdaten der Lernmodule.

**Keywords:** Autorensystem, WBT, Barrierefreiheit, Learning Analytics, agiles Arbeiten, Usability

### 1 Umsetzung der Barrierefreiheit

Der Fokus des neuen Release 4.6 liegt auf der Barrierefreiheit der Lernmodule. Die Grundfunktionen der LernBar-Kurse sollen mit der Tastatur, alternativen Eingabemethoden oder einem Screenreader bedien- und navigierbar sowie das Design der Kurse für Personen mit Seheinschränkungen (bis zur Blindheit) nutzbar gemacht werden. In Kooperation mit dem BliZ<sup>2</sup> wurden über 50 User-Stories entwickelt, die die Barrierefreiheit von Kursen maßgeblich verbessern und eine Erfüllung der Kriterien der WCAG 2.1<sup>3</sup> näherbringen. Aufgrund der Komplexität einiger User-Stories (z.B. Drag&Drop-Fragetyp) konnten bisher noch nicht alle Anforderungen erfüllt werden. Eine Expert\*innen-Evaluierung zur Erfüllung der WCAG-Richtlinien sowie Tests mit Screenreadern und Tastatur-Navigation haben ergeben, dass LernBar-Kurse in der neuen Version für eine größere Zielgruppe von Personen in ihren Grundfunktionen nutzbar sind<sup>4</sup>.

---

<sup>1</sup> Goethe-Universität Frankfurt am Main, studiumdigitale, {voss | sacher | gattinger | weiss} @studiumdigitale.uni-frankfurt.de

<sup>2</sup> Zentrum für blinde und sehbehinderte Studierende der Technischen Hochschule Mittelhessen

<sup>3</sup> Die Web Content Accessibility Guidelines sind Richtlinien für barrierefreie Webinhalte

<sup>4</sup> Unter der Voraussetzung, dass Autor\*innen bei der Erstellung auf eine barrierefreie Gestaltung achten

## 2 Sammlung und Auswertung von Nutzerdaten

Zur Datenerfassung in Learning Analytics (LA) Prozessen bietet die LernBar, neben SCORM, die Möglichkeit, Nutzungsdaten in einem Learning Record Store im xAPI-Format aufzuzeichnen. Mit Hilfe dieser xAPI-Implementierung werden mehrere LA Projekte umgesetzt und durch Anpassungen und Erweiterungen stetig weiterentwickelt und beforscht ([WSS16], [SG20]). Auf Basis dieser, meist durch Lösung individueller Probleme erlangten Erkenntnisse, wurde die Implementierung der LernBar überarbeitet. Wohlüberlegte Erweiterungen der Metadaten in den einzelnen Statements sowie der Activity-Beschreibungen machen die Daten leichter verständlich und besser analysierbar. Ziel ist es, Autor\*innen eine automatische Auswertung der Nutzungsdaten sowie aussagekräftige Visualisierungen ihrer Kurse anzubieten. Ein Dashboard, welches Gebrauch von den neuen Datenstrukturen macht, wurde konzeptionell entworfen sowie prototypisch umgesetzt und soll in Zukunft den Autor\*innen zur Verfügung gestellt werden.

## 3 Fazit und Ausblick

Mit der LernBar produzierte Lernmodule sind nun mit assistierenden Technologien wie Screenreadern leichter zugänglich. Grundlegende Kriterien der WCAG werden erfüllt und die noch offenen User-Stories, wie bspw. die barrierearme Umsetzung des Drag&Drop-Fragetyps, werden in weiteren Entwicklungssprints sukzessiv weiterbearbeitet. Das Entwicklungsteam hat es sich nach der Implementierung zum Ziel gesetzt, LernBar-Kurse als WCAG-konform zertifizieren zu lassen. Die überarbeitete Speicherung von Nutzungsdaten soll Lehrenden ermöglichen, mit Hilfe eines LA-Dashboards, ihre Inhalte zu evaluieren und zu verbessern. Parallel wird daran gearbeitet, den "offline-first" Autor\*innenprozess im Windows-basierten LernBar Studio in eine Webapplikation zu überführen. Durch die Vereinigung mit einem neu konzipierten LernBar Portal soll eine barrierefreie Webplattform geschaffen werden, die umfangreiche Möglichkeiten zum Erstellen, Teilen, Bearbeiten, Veröffentlichen und Analysieren von Lernmodulen bietet.

## Literaturverzeichnis

- [WE18] Weiß, D.: Web-based Trainings im Kontext offener Lehr-/Lernprozesse: Neukonzeption, Implementierung und Szenarien, Dissertation, Universitätsbibliothek der Goethe-Universität, Frankfurt, 2018.
- [WSS16] Weiß, D.; Sacher, P.; Schiffner, D.: In-place content monitor to enhance and improve the authoring process of a web based training, ICERI2016 Proceedings, S. 892-899, 2016.
- [SG20] Sacher, P.; Gattinger, T.: Learning Activity Provider, INTED2020 Proceedings, S. 5645-5654, 2020.

## InfoBiTS - Informatische Bildung für Technikferne Seniorinnen und Senioren

Svenja Noichl<sup>1</sup>, Ulrik Schroeder<sup>1</sup>

**Abstract:** Digitale Kompetenzen werden auch für Seniorinnen und Senioren immer wichtiger. Damit insbesondere unerfahrene Nutzerinnen und Nutzer, sowie Personen mit (altersbedingten) Einschränkungen digitale Lernangebote zur Förderung digitaler Kompetenzen ohne größere Einschränkungen nutzen können, kombiniert InfoBiTS Lernmodule zur Förderung digitaler Kompetenzen sowie Grundlagen-Module zum Erlernen von u. a. Gesten und Personalisierung.

**Keywords:** Digitale Kompetenzen, Personalisierung, Technikferne Erwachsene

### 1 Konzept und Implementierung

In Zeiten voranschreitender Digitalisierung werden digitale Kompetenzen auch für Seniorinnen und Senioren immer wichtiger. Allein beim Gedanken an Geräte mit Touchscreen fällt auf, dass diese zunehmend in den Alltag integriert sind. Neben Smartphones und Tablets, die auch unter Seniorinnen und Senioren eine immer größere Verbreitung aufweisen, sind Touchscreens z. B. in Fahrkartenautomaten oder Bankautomaten zu finden. Gerade für (ältere) Nutzerinnen und Nutzer ohne Vorerfahrungen werden dadurch folgende Aspekte besonders relevant:

1. Individuelle Bedürfnisse und (altersbedingte) Einschränkungen der Nutzerinnen und Nutzer müssen berücksichtigt werden, um die Nutzung digitaler Geräte im Allgemeinen und digitaler Lernangebote im Speziellen nicht zu erschweren.
2. Grundlegende Kenntnisse zur Nutzung solcher Geräte, wie z. B. Touch-Gesten und Eingabemöglichkeiten, sind unverzichtbare Fähigkeiten insbesondere zur Nutzung von Geräten mit Touchscreen.
3. Digitale Kompetenzen müssen gefördert werden, um eine aktive und selbständige Teilhabe am digitalen Alltag zu ermöglichen.

InfoBiTS ist eine Lernapplikation, die diese Aspekte kombiniert. Im Fokus stehen vier Lernmodule zur Vermittlung von grundlegenden Informatikkenntnissen zur Förderung digitaler Kompetenzen [NS19a, NS19b]. In diesen Modulen liegt neben der Wissensvermittlung ein besonderer Fokus auf integrierten, interaktiven Übungsaufgaben. Insbesondere bei den Übungen ist ein sicherer Umgang mit gängigen Touch-Gesten unverzichtbar. Daher bietet InfoBiTS zusätzliche Grundlagen-Module, wie z. B. ein Modul zum Erlernen der Gesten Antippen, Scrollen, Wischen, Zoomen und Drag & Drop.

---

<sup>1</sup> RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, {noichl, schroeder}@informatik.rwth-aachen.de

Um eventuell auftretenden (altersbedingten) Einschränkungen entgegenzuwirken verfügt InfoBiTS zudem über ein Personalisierungsmodul. Dieses ermöglicht eine individualisierte Anpassung von Schriftgröße, Buttongröße sowie der Textposition im Button, an die eigenen Bedürfnisse und Nutzungsgewohnheiten. Wichtig hierbei ist es, dass sich gerade Smartphone- und Tablet-Neulinge nicht durch die zum Teil komplexen und von Hersteller und Betriebssystem abhängigen Einstellungsmenüs der Geräte suchen müssen. Besondere Bedeutung haben hier die Einstellungen von Schriftgröße, Buttongröße sowie die Positionierung der Schrift innerhalb des Buttons. Durch die Anpassung dieser Parameter wird den Nutzerinnen und Nutzern ein Treffen der Schaltflächen vereinfacht [NRS18]. So werden mögliche Nutzungsprobleme minimiert und der Fokus kann im Folgenden auf den Lerninhalten der Lernapplikationen liegen.

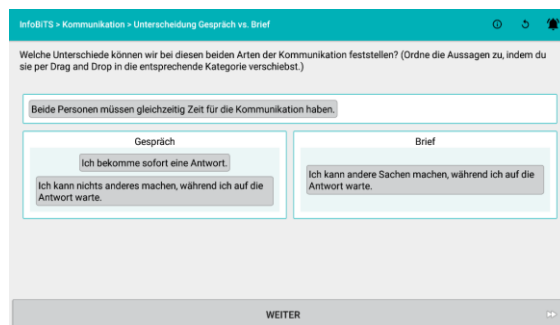


Abb. 1: exemplarische, interaktive Aufgaben

Dieser Beitrag und die Demonstration zeigt die Umsetzung einer Lernapplikation für die Zielgruppe der Seniorinnen und Senioren zur Förderung digitaler Kompetenzen, die neben inhaltlichen Modulen auch Grundlagen der Interaktion mit Touchscreen-Geräten behandelt, sowie eine individuelle Anpassung der Nutzungsoberfläche bietet. Eine Evaluation mit der Zielgruppe, welche digitalen Kompetenzen in welchem Maße durch die Nutzung der App gefördert werden, befindet sich in der ersten Durchführungsphase.

## Literaturverzeichnis

- [NRS18] Noichl, S.; Röpke, R.; Schroeder, U.: Adaptive Buttons für zielgruppengerechtes App Design. In: Dachzelt, R. & Weber, G. (Hrsg.): Mensch und Computer 2018 - Workshopband. Gesellschaft für Informatik e.V., Bonn, 2018.
- [NS19a] Noichl, S.; Schroeder, U.: Informatik Grundbildung im Alter – Ein eLearning Konzept. In: Pinkwart, N. & Konert, J. (Hrsg.), DELFI 2019. Gesellschaft für Informatik e.V., Bonn, S. 309-310, 2019.
- [NS19b] Noichl, S.; Schroeder, U.: Zu alt für Informatik?: Seniorinnen und Senioren erobern die digitale Welt. In: David, K., Geihs, K., Lange, M. & Stumme, G. (Hrsg.), INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft. Gesellschaft für Informatik e.V., Bonn, S. 685-698, 2019.

## Ein Webbasiertes Tool zur Konvertierung von Bildungsmaterialien in OER

Lubna Ali<sup>1</sup>, Patrick Aufdermauer<sup>2</sup>, René Röpke<sup>1</sup>, Ulrik Schroeder<sup>1</sup>

**Abstract:** Open Educational Resources (OER) sind mittels offener Lizenzmodelle lizenzierte, freie Bildungsmaterialien. Die Verwendung offener Lizenzmodelle führt bei der Erstellung von OER zu rechtlichen Fragen und Unsicherheit. Zudem ist die Erstellung von OER mit einem hohen Aufwand verbunden. Es gibt zwar Werkzeuge, die helfen, OER von Grund auf zu erstellen, aber keine Werkzeuge, die darin unterstützen bereits erstellte Lehrmaterialien in OER umzuwandeln. Die Herausforderung liegt in der Aufbereitung der Materialien und das Ergänzen wichtiger Lizenzinformationen. Dieser Beitrag stellt ein webbasiertes Tool vor, das mit vergleichsweise geringem Aufwand aus vorhandenen Lehrmaterialien OER-konformes Dokumente erstellt.

**Keywords:** OER, Lehrmaterialien, Lizenzierung, Konvertierung, Automatisierung

### Konzept und Implementierung

Damit Materialien als OER gelten, müssen fünf Bedingungen für Offenheit erfüllt werden [Hi10]: (1) *Retain*: Erlaubnis sich Kopien anzulegen, (2) *Reuse*: Erlaubnis den Inhalt zu benutzen, zu verwenden und jemandem zu zeigen, (3) *Revise*: Erlaubnis den Inhalt anzupassen oder zu verändern, (4) *Remix*: Erlaubnis den Inhalt (oder den bereits veränderten Inhalt) mit anderen Inhalten neu zu kombinieren, um einen neuen Inhalt zu erzeugen und (5) *Redistribute*: Erlaubnis den originalen Inhalt zu verändern oder neu zusammengestellten Inhalt zugänglich zu machen.

Die Erfüllung dieser fünf Bedingungen stellt Lehrende in der Konvertierung ihrer existierenden Lehrmaterialien vor eine große Herausforderung. Zur Unterstützung wird ein webbasiertes Analysetool entwickelt [A119], welches eine Datei einliest und Bildelemente sowie mögliche Metadaten aufbereitet. Nicht-OER-konforme Abbildungen sollen hierbei durch geeignete Alternativen mit offener Lizenzierung ersetzt werden.

Das Tool soll dabei teilautomatisiert aus bereits bestehendem Lehrmaterial ein OER-konformes Ergebnisdokument erstellen. Dazu muss das Tool in der Lage sein, automatisiert auf die verwendeten Bilder zuzugreifen und diese bei Bedarf auszutauschen. In der aktuellen Version kann das Tool eine Präsentationsdatei für Microsoft PowerPoint einlesen und die Bilder extrahieren.

---

<sup>1</sup> RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, {ali, röpke, schroeder}@informatik.rwth-aachen.de

<sup>2</sup> RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, patrick.aufdermauer@rwth-aachen.de

Anschließend sollen alle Bilder innerhalb des Dokuments geprüft und bei Bedarf durch offenzensiertes Bildmaterial inkl. Lizenzhinweis ersetzt werden.

Das offenzensierte Bildmaterial zum Ersetzen wird durch eine Anbindung zu unterschiedlichen OER-Portalen und Datenbanken im Tool eingebunden. Eine Suchfunktionalität unterstützt das Auffinden passender Bilder zum Ersetzen. Durch die Anbindung können Metadaten und insb. Lizenzinformationen von externen Quellen abgerufen und in die zu konvertierende Datei eingebunden werden. Die Folien werden dabei manipuliert und zum abschließend wird eine neue Folie mit Lizenzinformationen des Gesamtdokuments angehängt.

Das webbasierte Analysetool besteht aus zwei Komponenten: Das Frontend bietet die Benutzerschnittstelle, die das Einlesen, die Konvertierung und den abschließenden Download der Lehrmaterialien realisiert. Das Backend erfasst hierbei, welche konkreten Substitutionsoperationen stattfinden, d.h. welche ursprünglichen Bilder in der Präsentation durch neue Bilder aus den zur Verfügung stehenden externen Quellen ersetzt werden. Später kann auf diese Informationen zurückgegriffen werden um zukünftiges Material mit gleichen Bildern (teil-)automatisiert zu konvertieren.

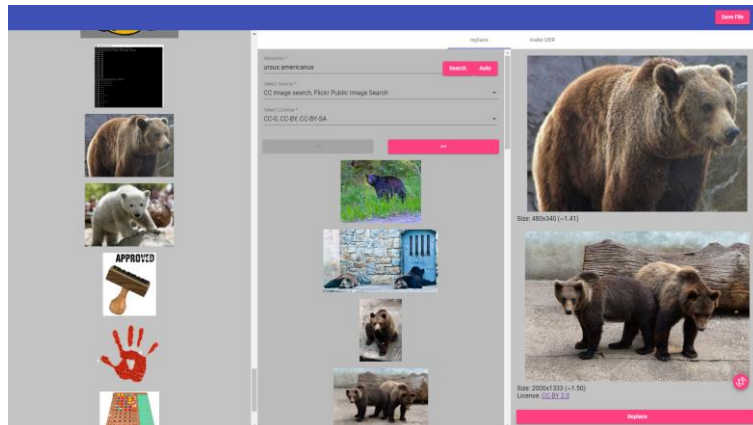


Abb. 1: Links sind die Bilder in der Präsentation zu sehen, in der Mitte die einer externen Bildersuchmaschine. Rechts ist das alte Bild (oben) und das neue Bild (unten) zu sehen

## Literaturverzeichnis

- [Hi10] Hilton III, J. et al.: The four 'r's of openness and alms analysis: Frameworks for open educational resources, *Open Learning*, vol. 25, pp. 37–44, 02 2010.
- [Al19] Ali L.; Schroeder, U.: A tool to enhance the utilization of OER in higher education, *Proc. 11th International Conference on Education and New Learning Technologies (EDULEARN19)*, pp. 7321-7325, 07 2019.



## An Environment-Triggered Augmented-Reality Application for Learning Case Grammar

Fiona Draxler<sup>1</sup>, Elena Wallwitz<sup>1</sup>, Albrecht Schmidt<sup>1</sup>, Lewis L. Chuang<sup>1</sup>

**Abstract:** We present a handheld Augmented-Reality app that enables learners of German to study case grammar using real objects in their surroundings. The system can easily be integrated into the everyday life of busy learners and provides them with a means to study on their own. Specifically, the app detects objects in the learners' surroundings and determines their spatial relationship. It then automatically generates quizzes to test dative and accusative cases. The app provides an example of how structural concepts of languages can be taught with AR.

**Keywords:** Augmented Reality; Case Grammar; Language Learning; Ubiquitous Learning

### 1 Introduction

Augmented Reality (AR) is considered a promising technology for supporting learning because AR learning systems can provide timely context-relevant content in a real-world setting [SK18]. Moreover, AR is suitable for visualizing spatial or temporal concepts and for learners to actively engage with learning content [BD12]. For example, in language learning, ARbis Pictus used AR goggles to show foreign-language labels [Ib18]. This led to more successful learning than a flashcard-based learning approach. Grammatical aspects of language were enhanced in a previous project of ours [Dr20]: we introduced a handheld AR app for studying grammatical gender and the dative case in German. We used markers to detect objects and their spatial relationship in the learners' surroundings to construct sentence quizzes. In the current project, we extend this approach with automated object detection and additional variation in quizzes.

### 2 Concept and Implementation

In this demo, we utilize the affordance of AR for visualizing spatial relationships and apply it to challenging aspects of the German language: grammatical gender and case grammar. Specifically, we automatically identify objects in the learners' surroundings and use them to create multiple-choice quizzes: we build a sentence describing the relative position of two objects where learners need to select the article that (1) has the grammatical gender of the second object and (2) is correctly inflected in dative or accusative case. The interaction with the learners' surroundings increases contextual relevance and most likely also the motivation for learning.

---

<sup>1</sup> LMU Munich, Human-Centered Ubiquitous Media, Frauenlobstr. 7a, 80337 München, first.last@ifi.lmu.de

The app was built for Android devices in order to reach a wide audience. It uses the back camera to continuously scan and identify objects learners point the device at. It shows a live camera preview and augments it with labels of detected objects (see Figure 1). The camera image is processed in a Tensorflow Lite Object detection<sup>2</sup> workflow using a model trained on the Common Objects in Context (COCO) data set<sup>3</sup>, which contains 80 object classes. Labels are shown when the confidence threshold for a detected object is reached. As soon as a learner has selected two labels (i.e., objects) on the screen, their relative position is computed and a quiz is shown at the bottom of the screen (see Figure 2).

Thus, to summarize, with our novel app, learners can practice challenging aspects of the German language whenever and wherever it suits them. With slight changes, such an app would also be suitable for studying similar problems, such as cases in other languages.



Fig. 1: Screenshot AR Case Grammar Learner

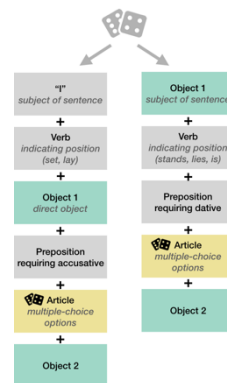


Fig. 2: Quiz Generation Process

## Literaturverzeichnis

- [BD12] Billinghamurst, M., & Duenser, A.: Augmented Reality in the Classroom. *Computer* 45(7), pp. 56–63, 2012.
- [Dr20] Draxler, F., Labrie, A., Schmidt, A., & Chuang L. L.: Augmented Reality to Enable Users in Learning Case Grammar from Their Real-World Interactions. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems - CHI '20*. ACM Press, pp. 1–12, 2020.
- [Ib18] Ibrahim, A., Huynh, B., Downey, J., Höllerer, T., Chun, D., & O'donovan, J.: ARbis Pictus: A Study of Vocabulary Learning with Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 24(11), pp. 2867–2874, 2018.

<sup>2</sup> [https://www.tensorflow.org/lite/models/object\\_detection/overview](https://www.tensorflow.org/lite/models/object_detection/overview)

<sup>3</sup> <http://cocodataset.org/>

## Verzeichnis der Autorinnen und Autoren

Akao, Kensuke.....	367	Franzkeit, Janna .....	365
Ali, Lubna .....	387	Funk, Johannes.....	25
Aufdermauer, Patrick.....	387	Funke, Florian.....	307
Barbas, Helena .....	313	Fussan, Carsten .....	67
Bauer, Nicola H. ....	103	Gamst, Cornelia .....	377
Bauer, Timon .....	295	Gattinger, Thorsten .....	121, 383
Bejan, Alexander.....	79	Giorgashvili, Tornike .....	169
Bender, Esther.....	313	Girvan, Carina.....	247
Bergerbusch, Timo.....	133	Greiff, Samuel.....	21
Bergner, Nadine .....	97	Greubel, Andre.....	351
Bernloehr, Annette .....	103	Günther, Franziska.....	199
Bertelsmeier, Birgit.....	211	Haake, Joerg M. ....	145
Blattgerste, Jonas .....	103	Hahn, Kathrin.....	265
Blömer, Linda .....	343	Hainke, Carolin.....	355
Brix, Orlando .....	127	Haller, David.....	223
Chuang, Lewis L.....	389	Hamann, Fabian .....	313
Claus, Sebastian .....	361	Hanesova, Veronika .....	67
Dahm, Markus .....	259	Hassan, Ahmed .....	379
Damnik, Gregor .....	97	Heinemann, Birte .....	331
Dehne, Julian .....	319	Hennecke, Martin.....	351
Drachsler, Hendrik.....	381	Heuts, Alexander.....	277
Draxler, Fiona .....	389	Hofmann, Jens .....	49
Dufentester, Christian .....	377	Hofmann, Sven .....	295, 307
Durski, Sara .....	363	Hoppe, Uwe .....	343
Ehlenz, Matthias .....	331	Höppner, Frank .....	217
Eitemüller, Carolin.....	109	Horn, Florian.....	121
Faeskorn-Woyke, Heide .....	211	Hörnlein, Stefanie .....	353
Falke, Tobias.....	193	Ionica, Lavinia .....	377

Jäger, Kathrin.....	205, 373	Merten, Maik .....	229
Joswig, Matthias .....	103	Meyer, Leonard.....	55
Judel, Sven.....	133	Müller, Wolfgang.....	363
Karolyi, Heike.....	145	Neumann, Jörg .....	199
Kastner, Marvin .....	365	Noichl, Svenja.....	385
Kersten, Stefanie .....	295	Osmers, Niklas .....	37
Kiesler, Natalie .....	187	Pfänder, Peter.....	337
Köchling, Alina.....	369	Pfeiffer, Thies .....	
Köhler, Marius .....	259	.....	55, 103, 355, 359, 371
König, Peter .....	79	Pinkwart, Niels.....	
Kunze, Christophe.....	79	.....	265, 277, 289, 361, 379
Kunzendorf, Martina.....	103	Plotzky, Christian.....	79
Lainé, Anna.....	365	Prilla, Michael.....	37
Lehmann, Marcel .....	295	Quandt, Joachim .....	265
Lenz, Richard.....	223	Riazy, Shirin .....	301
Leonhardt, Thiemo.....	97	Rittberger, Marc .....	381
Leßner, Tobias .....	73	Ritter, Steffen.....	151
Lewa, Carmen .....	103	Röpke, René.....	331, 387
Lindwedel-Reime, Ulrike.....	79	Rose, Jennifer.....	259
Lorenz, Robert .....	199	Röbler, Christian .....	151
Lucke, Ulrike .....	319	Rüdian, Sylvio .....	265
Lukarov, Vlatko .....	331	Rudolph, Theresa .....	351
Luksch, Kristina.....	103	Rustemeier, Linda .....	157
Ma, Lihong.....	145	Sacher, Patrick .....	121, 383
Mah, Dana-Kristin .....	377	Schäfer, Thorsten .....	103
Malotky, Nikolaj Troels Graf von..		Schiffner, Daniel.....	121
.....	91	Schmidt, Albrecht .....	389
Martens, Alke.....	91	Schmidt, Ludger.....	25
Massler, Ute .....	363	Schneider, Jan .....	381
Mateen, Saba.....	157	Schober, Susann.....	127
Meinel, Christoph .....	235	Schöneburg-Lehnert, Silvia .....	295

Schröder, Nadine.....	337	Suchodoletz, Dirk von .....	151
Schroeder, Ulrik.....		Tehreem, Yusra.....	359
.....	133, 331, 385, 387	Teusner, Ralf.....	235
Seidel, Niels .....	145	Thüring, Annett.....	205
Semm, Arlett.....	181	Träger, Marco .....	301
Serth, Sebastian.....	235	Trauten, Florian .....	109
Seufert, Sabine .....	19	Umlauft, Annegret .....	199
Simbeck, Katharina.....	301	Voigt, Christin .....	343
Sitzmann, Daniel.....	313	Voß-Nakkour, Sarah .....	383
Slotosch, Sven.....	151	Wahl, Andreas .....	223
Söbke, Heinrich .....	353	Wallwitz, Elena.....	389
Soll, Marcus .....	313	Wehking, Florian .....	353
Spieler, Bernadette.....	247	Wehner, Marius .....	369
Spierling, Ulrike.....	181	Weiß, David .....	383
Stalljohann, Patrick .....	229	Wilhelm-Weidner, Arno .....	357
Steinke, Leena.....	127	Wolf, Mario .....	353
Strickroth, Sven .....	127, 319	Wollny, Sebastian .....	381
Striewe, Michael .....	109	Wöstenfeld, Robert .....	301
Strohschein, Jan .....	211		

