

Ilka Koppel

Entwicklung einer Online-Diagnostik für die Alphabetisierung

Eine Design-Based Research-Studie



Springer VS

Entwicklung einer Online-Diagnostik für die Alphabetisierung

Ilka Koppel

Entwicklung einer Online-Diagnostik für die Alphabetisierung

Eine Design-Based Research-Studie

Ilka Koppel
Bremen, Deutschland

Dissertation Universität Bremen, 2015

„Iterative Entwicklung und Evaluation einer pädagogischen Online-Diagnostik für funktionale Analphabeten und Analphabetinnen. Anwendungsorientierte und theoretische Folgerungen für eine zielgruppenspezifische gebrauchstaugliche Gestaltung“ wurde im Jahr 2015 an der Universität Bremen verteidigt und im Jahr 2016 unter dem Buchtitel „Entwicklung einer Online-Diagnostik für die Alphabetisierung. Eine Design-Based Research-Studie“ von Springer VS-Research veröffentlicht.

ISBN 978-3-658-15768-5

ISBN 978-3-658-15769-2 (eBook)

DOI 10.1007/978-3-658-15769-2

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden GmbH 2017

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen.

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist Teil von Springer Nature

Die eingetragene Gesellschaft ist Springer Fachmedien Wiesbaden GmbH

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Inhalt

Tabellenverzeichnis	11
Abbildungsverzeichnis	15
Abkürzungsverzeichnis.....	19
Einleitung	21
1.1 Motivation und Zielsetzung der Arbeit.....	21
1.2 Aufbau der Arbeit	25
Phase I – Problemanalyse.....	27
2 Theoretische Herleitung.....	29
2.1 Diagnostische Grundlagen.....	29
2.1.1 Pädagogische Diagnostik	29
2.1.2 Förderdiagnostik.....	33
2.1.3 Zusammenfassender Vergleich und Akzeptanz diagnostischer Verfahren	34
2.2 Computerbasierte Diagnostik.....	36
2.2.1 Computerbasierte Diagnostik – Definition und Abgrenzung.....	36
2.2.2 Rich E-Assessment	40
2.2.3 Vorteile computerbasierter Diagnostik.....	41
2.2.4 Potenziale computerbasierter Diagnostik für die Förderdiagnostik.....	42
2.2.5 Beispiele computerbasierter Diagnostik.....	46
2.2.6 Innovative Itemformate	52
2.2.6.1 Dimensionen innovativer Itemformate	54
2.2.6.2 Taxonomie innovativer Itemformate	63
2.3 Zielgruppenspezifische Voraussetzungen	66
2.3.1 Literalität.....	66
2.3.2 Funktionaler Analphabetismus.....	67
2.3.3 Funktionaler Analphabetismus in Deutschland, England und Frankreich.....	70

2.3.4	Herausforderungen in der Alphabetisierung in Deutschland.....	73
2.3.5	Informations- und Computertechnologie (ICT)-Literacy	75
2.3.6	Computerbasierte Lern- und Diagnoseinstrumente im Grundbildungsbereich Erwachsener	78
2.4	Cognitive Load Theory und Cognitive Theory of Multimedia Learning.....	80
2.4.1	Grundidee und Annahmen der Cognitive Load Theory	80
2.4.2	Cognitive Theory of Multimedia Learning.....	84
2.4.3	Messmöglichkeiten des Cognitive Load.....	87
2.5	Exkurs: Äquivalenzproblematik	89
2.6	Usability	91
2.6.1	Usability – eine Begriffsan- und -einordnung	91
2.6.2	User Experience (UE)	96
2.6.3	Kognitive und physiologische Voraussetzungen	97
2.6.3.1	Aufmerksamkeit.....	98
2.6.3.2	Wahrnehmung.....	99
2.6.3.3	Wissensrepräsentation und Mentale Modelle.....	101
2.6.3.4	Physiologische Voraussetzungen	103
2.6.4	Allgemeine Usability-Anforderungen und Heuristiken ..	104
2.6.4.1	Usability-Heuristiken.....	106
2.6.4.2	Accessibility und die Web Content Accessibility Guidelines (WCAG)	116
2.6.4.3	Heuristiken – Grenzen der Anwendbarkeit.....	120
2.6.4.4	Zusammenfassung der allgemeinen Usability-Anforderungen und der WCAG.....	122
2.6.5	Vorläufige Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen	125
2.7	Evaluation von Usability	127
2.7.1	Formal -Analytischer Messansatz.....	129
2.7.2	Produktzentrierter Messansatz	129
2.7.3	Interaktionszentrierter Messansatz.....	132
2.7.4	Benutzer-/Benutzerinnenorientierte Messansatz.....	133
2.7.5	Usability-Tests in Laboratories.....	134
2.7.6	Planung und Durchführung von Usability-Evaluation	135

2.8 Zusammenfassung der theoretischen Herleitung	137
PHASE II – Konzeption	141
3 Konzeption des Forschungsrahmens - Design-Based Research (DBR)	143
3.1 Ziele des DBR-Ansatzes	144
3.1.1 Merkmale von DBR.....	149
3.1.2 Ablauf von DBR.....	150
3.1.3 Forschungsstrategisches und -methodisches Vorgehen.....	154
3.1.4 Der DBR-Ansatz im wissenschaftlichen Kontext.....	155
3.1.4.1 Verortung des DBR im Kontext der Wissenschaftstraditionen	155
3.1.4.2 DBR in Abgrenzung zu weiteren Ansätzen in der Lehr-Lernforschung	158
3.1.5 DBR in der Anwendung	160
3.1.6 Kritik am DBR-Ansatz	164
4 Testumgebung.....	169
4.1 Projektkontext: lea. – Literalitätsentwicklung von Arbeitskräften	169
4.2 Entwicklung der Online Testumgebung otu.lea	172
4.2.1 Herausforderungen für die Entwicklung	174
4.2.2 Das Konzept von otu.lea.....	175
PHASE III – Konzeptevaluation.....	179
5 Das Forschungsdesign	181
5.1 Entwicklung der Forschungsfrage und des Forschungsdesigns.....	181
5.2 Ziele des Forschungsvorhabens	183
5.3 Ablauf des dritten Zyklus` und Forschungsdesign.....	185
5.4 Forschungsstrategisches und -methodisches Vorgehen.....	190
5.5 Kritik am DBR-Ansatz im Kontext der Fragestellung.....	192

6	Zyklus 1: Paper Prototyping	195
6.1	Zyklus 1: Gestaltung	195
6.2	Zyklus 1: Durchführung	199
6.3	Zyklus 1: Analyse	200
6.4	Zyklus 1: Re-Design	204
6.5	Reflexion des Paper Prototyping-Experiments	205
7	Zyklus 2: Usability-Studie 2012	207
7.1	Zyklus 2: Gestaltung	207
7.2	Zyklus 2: Durchführung	210
7.3	Zyklus 2: Analyse	210
7.4	Zyklus 2: Re-Design	217
7.5	Reflexion der Usability-Studie 2012	223
7.6	Zusammenfassung und Reflexion hinsichtlich des dritten Zyklus	224
7.7	Weitere Entwicklungsergebnisse der Online- Testumgebung	225
7.7.1	Die Technologieplattform von otu.lea	225
7.7.2	Systemgenerierte Auswertung mit otu.lea	227
7.7.3	Rich E-Assessment und Itemformate in otu.lea	227
8	Zyklus 3: Usability-Studie 2013	235
8.1	Zyklus 3: Gestaltung	235
8.1.1	Entwicklung und Definition des Forschungsgegenstands	235
8.1.2	Formulierung der Forschungsfragen und Hypothesen	236
8.1.2.1	Operationalisierung der Forschungsfragen und der Hypothesen	240
8.1.2.2	Methodische Vorüberlegungen	246
8.2	Zyklus 3: Durchführung – Methodenwahl und Untersuchungskontext	249
8.2.1	Forschungsdesign und Ablauf der Hauptuntersuchung	250

8.2.2	Die Erhebungsinstrumente.....	252
8.2.2.1	Leitfadengestützte Interviews.....	252
8.2.2.2	Videoaufzeichnung.....	259
8.2.2.3	Befragungen.....	260
8.2.2.4	Data-Logging mit otu.lea	262
8.2.3	Zugang zur Stichprobe.....	262
8.2.4	Auswertungsmethoden	262
8.2.5	Einhaltung der Gütekriterien.....	266
8.3	Zyklus 3: Analyse – Darstellung der Ergebnisse.....	268
8.3.1	Demographische Angaben, Bearbeitungszeit und Abbruchquote.....	268
8.3.2	Allgemeine deskriptive Ergebnisdarstellung.....	268
8.3.2.1	<i>Allgemeine Ergebnisdarstellung zur Performanz</i>	<i>269</i>
8.3.2.2	<i>Allgemeine Ergebnisse der Videoanalyse.....</i>	<i>272</i>
8.3.2.3	<i>Allgemeine Ergebnisdarstellung zum User Experience Questionnaire.....</i>	<i>275</i>
8.3.2.4	<i>Allgemeine Ergebnisse der Interviews</i>	<i>276</i>
8.3.2.5	<i>Allgemeine Ergebnisse zur Computererfahrung und - kompetenz.....</i>	<i>280</i>
8.3.3	Ergebnisdarstellung zu den Usability-Heuristiken.....	284
8.3.4	Zusammenfassung und Diskussion der Ergebnisse zu den Usability-Heuristiken.....	315
8.3.5	Ergebnisdarstellung zur Computererfahrung und - kompetenz.....	317
8.3.6	Diskussion der Ergebnisse zur Computererfahrung und -kompetenz.....	323
8.3.7	Darstellung der Ergebnisse zum Extraneous Load.....	326
8.3.8	Diskussion der Ergebnisse zum Extraneous Load	334
8.4	Zusammenfassung der Ergebnisse.....	340
8.5	Zyklus 3: Re-Design.....	342
8.6	Zusammenfassende Diskussion und Schlussfolgerung	347

Phase IV – Reflexion.....	355
9 Reflexion der Erhebungsmethoden und -instrumente.....	357
9.1 Reflexion des Forschungsrahmens DBR.....	357
9.2 Kritische Reflexion der Erhebungsmethoden und -instrumente	359
9.3 Stichprobenauswahl	363
10 Forschungsdesiderata und Ausblick.....	365
Literatur.....	371
Anhang	405

Tabellenverzeichnis

<i>Tabelle 1:</i>	<i>Dimensionsausprägung selektive Diagnostik und Förderdiagnostik.....</i>	<i>35</i>
<i>Tabelle 2:</i>	<i>Taxonomie innovativer Itemformate).....</i>	<i>64</i>
<i>Tabelle 3:</i>	<i>Beruflicher Status funktionaler Analphabetismus und fehlerhaftes Schreiben).....</i>	<i>71</i>
<i>Tabelle 4:</i>	<i>Ausstattung der Haushalte mit Computern 2012.....</i>	<i>76</i>
<i>Tabelle 5:</i>	<i>Angenommener Zusammenhang zwischen der Beanspruchung des Cognitive Load und Usability.....</i>	<i>87</i>
<i>Tabelle 6:</i>	<i>Vorläufige Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen.....</i>	<i>126</i>
<i>Tabelle 7:</i>	<i>Charakteristiken von DBR.....</i>	<i>150</i>
<i>Tabelle 8:</i>	<i>Arbeitsschritte im Teilprojekt zur Entwicklung von otu.lea</i>	<i>173</i>
<i>Tabelle 9:</i>	<i>Konkretisierte Darstellung des Forschungsdesigns Ex-Post.....</i>	<i>187</i>
<i>Tabelle 10:</i>	<i>DBR-Zyklen und Schritte der Usability-Evaluation.....</i>	<i>189</i>
<i>Tabelle 11:</i>	<i>Übersicht Paper Prototyping-Experiment 2010.....</i>	<i>198</i>
<i>Tabelle 12:</i>	<i>Versuchsanordnung Paper Prototyping 2010.....</i>	<i>198</i>
<i>Tabelle 13:</i>	<i>Ergebnisse des Paper Prototyping 2010.....</i>	<i>202</i>
<i>Tabelle 14:</i>	<i>Modifikationen von otu.lea auf der Grundlage des ersten Zyklus.....</i>	<i>205</i>
<i>Tabelle 15:</i>	<i>Übersicht Usability-Studie 2012</i>	<i>209</i>
<i>Tabelle 16:</i>	<i>Ergebnisse der Usability-Studie 2012</i>	<i>213</i>
<i>Tabelle 17:</i>	<i>Modifikationen von otu.lea auf der Grundlage des zweiten Zyklus</i>	<i>217</i>

<i>Tabelle 18:</i>	<i>Zusammenfassung der ersten zwei Evaluationszyklen.....</i>	<i>224</i>
<i>Tabelle 19:</i>	<i>Kategorisierung der otu.lea.-Items</i>	<i>230</i>
<i>Tabelle 20:</i>	<i>Experimentelle Bedingungen Usability-Studie 2013.....</i>	<i>251</i>
<i>Tabelle 21:</i>	<i>Übersicht der Datenerhebungen des dritten Evaluationszyklus</i>	<i>263</i>
<i>Tabelle 22:</i>	<i>Experimentelles Design /Versuchsplan Ex-Post.....</i>	<i>269</i>
<i>Tabelle 23:</i>	<i>Summenwerte format-, niveau- und dimensionsbezogen.....</i>	<i>271</i>
<i>Tabelle 24:</i>	<i>Relative Performanzwerte format-, niveau- und dimensionsbezogen</i>	<i>272</i>
<i>Tabelle 25:</i>	<i>Skalenkonsistenz für die niveau- und dimensionsbezogenen Itemformate.....</i>	<i>272</i>
<i>Tabelle 26:</i>	<i>Auswertungsleitfaden Videoanalyse.....</i>	<i>273</i>
<i>Tabelle 27:</i>	<i>Quantitative Auswertung der Videobeobachtungen.....</i>	<i>274</i>
<i>Tabelle 28:</i>	<i>Ergebnisse des UEQ für die Skala Attraktivität.....</i>	<i>275</i>
<i>Tabelle 29:</i>	<i>Ergebnisse des UEQ für die Skala Stimulation.....</i>	<i>276</i>
<i>Tabelle 30:</i>	<i>Auswertungsleitfaden für die fokussierten Interviews mit Kursteilnehmenden</i>	<i>277</i>
<i>Tabelle 31:</i>	<i>Ergebnisse der Reliabilitätsprüfung</i>	<i>278</i>
<i>Tabelle 32:</i>	<i>Auswertungsleitfaden für das Experten-/ Expertinnengespräch mit Alphabetisierungskräften.....</i>	<i>279</i>
<i>Tabelle 33:</i>	<i>Internetnutzung.....</i>	<i>282</i>
<i>Tabelle 34:</i>	<i>Nutzung der Hilfefunktion</i>	<i>289</i>
<i>Tabelle 35:</i>	<i>Unterstützung beim Einloggen</i>	<i>292</i>
<i>Tabelle 36:</i>	<i>Erfüllung der Heuristiken und Modifikationsbedarfe für otu.lea.....</i>	<i>316</i>
<i>Tabelle 37:</i>	<i>Computererfahrung und Anzahl der eingeforderten Unterstützung.....</i>	<i>318</i>

<i>Tabelle 38:</i>	<i>Computererfahrung und Unterstützung gesamt.....</i>	<i>318</i>
<i>Tabelle 39:</i>	<i>Computererfahrung und relative formatbezogene Performanzwerte für Durchlauf 1.....</i>	<i>321</i>
<i>Tabelle 40:</i>	<i>Computererfahrung und relative formatbezogene Performanzwerte für Durchlauf 1 und 2.....</i>	<i>322</i>
<i>Tabelle 41:</i>	<i>Verteilung zwischen experimenteller Bedingungen (Cue) und Computererfahrung.....</i>	<i>329</i>
<i>Tabelle 42:</i>	<i>Beobachtung reduzierender und auslastender Faktoren für den Cognitive Load.....</i>	<i>338</i>
<i>Tabelle 43:</i>	<i>Übersicht der eingesetzten Erhebungsmethoden und Messansätze.....</i>	<i>359</i>

Abbildungsverzeichnis

Abbildung 1:	Formen des technologiebasierten Assessments.....	38
Abbildung 2:	Arbeitsgedächtnis und Cognitive Load.....	82
Abbildung 3:	Usability in Beziehung zu angrenzenden/verwandten Begriffen.....	95
Abbildung 4:	Grafische Darstellung der vier Gestaltgesetze	100
Abbildung 5:	Gestaltungs-/Anwendungsrahmen für DIN EN ISO 9241	106
Abbildung 6:	Grundsätze der Usability.....	108
Abbildung 7:	Wirkungszusammenhänge basierend auf der theoretischen Herleitung.....	138
Abbildung 8:	DBR-Ebenen, -Prozess und -Ablauf.....	145
Abbildung 9:	Ziele von DBR	147
Abbildung 10:	DBR-Prozess nach Reeves.....	151
Abbildung 11:	Schematischer Ablauf DBR-Zyklen.....	154
Abbildung 12:	Ablauf otu.lea	176
Abbildung 13:	Forschungsdesign Ex-Ante - Version 1.....	181
Abbildung 14:	Forschungsablauf Ex-Ante - Version 2.....	182
Abbildung 15:	Ziele des Forschungsvorhabens in Anlehnung an den DBR-Prozess.....	184
Abbildung 16:	Schematischer Ablauf des DBR-Forschungsprozesses Ex-Post	185
Abbildung 17:	Paper Prototyping mit einem Papierprototyp von otu.lea, Sequenz 1	199
Abbildung 18:	Paper Prototyping mit einem Papierprototyp von otu.lea, Sequenz 2	200

<i>Abbildung 19: Exemplarischer Ausschnitt aus dem Usability-Test 2012</i>	<i>211</i>
<i>Abbildung 20: Einstiegsseite mit Videotutorial von otu.lea</i>	<i>218</i>
<i>Abbildung 21: Anmeldung und Registrierung.....</i>	<i>219</i>
<i>Abbildung 22: Dimensionsauswahl.....</i>	<i>219</i>
<i>Abbildung 23: Selbsteinschätzung.....</i>	<i>220</i>
<i>Abbildung 24: Itembeispiel aus der Dimension Lesen.....</i>	<i>221</i>
<i>Abbildung 25: Kontextsensitive Hilfefunktion</i>	<i>222</i>
<i>Abbildung 26: Ende der Diagnostik.....</i>	<i>222</i>
<i>Abbildung 27: lea.-Framework.....</i>	<i>226</i>
<i>Abbildung 28: Wirkungszusammenhänge und Hypothesen</i>	<i>240</i>
<i>Abbildung 29: Forschungsstrategisches Vorgehen des dritten Evaluationszyklus</i>	<i>250</i>
<i>Abbildung 30: Interviewleitfaden fokussiertes Interview Kursteilnehmende Usability-Studie 2013.....</i>	<i>256</i>
<i>Abbildung 31: Interviewleitfaden für das Experten-/ Expertinnengespräch mit Kursleitenden</i>	<i>259</i>
<i>Abbildung 32: Ablauf deduktiver Kategorienbildung</i>	<i>265</i>
<i>Abbildung 33: Positive Einstellung zur Computerarbeit</i>	<i>280</i>
<i>Abbildung 34: Balkendiagramm durchschnittliche Nutzungsdauer des Computers pro Woche.....</i>	<i>281</i>
<i>Abbildung 35: Durchschnittliche Gesamtbearbeitungszeit im Verhältnis zur durchschnittlichen eingeforderten Unterstützung</i>	<i>286</i>
<i>Abbildung 36: Vgl. Bearbeitungszeit und Unterstützungszeitpunkt (wenig Computererfahrung).....</i>	<i>287</i>
<i>Abbildung 37: Boxplot für Durchlauf 1: Cue und ausgewählte Aufgabenformate</i>	<i>327</i>

<i>Abbildung 38: Boxplot für Durchlauf 2: Cue und ausgewählte Aufgabenformate</i>	<i>328</i>
<i>Abbildung 39: Direkte und indirekte empirisch überprüfte Wirkungszusammenhänge</i>	<i>340</i>
<i>Abbildung 40: Rückmeldung für Teilnehmende</i>	<i>343</i>
<i>Abbildung 41: Report für Lernende "Das kann ich"</i>	<i>344</i>
<i>Abbildung 42: Report für Lernende „Das kann ich wenn ich noch ein wenig übe“</i>	<i>344</i>
<i>Abbildung 43: Ausschnitt aus dem Kursleitendenreport von otu.lea</i>	<i>345</i>
<i>Abbildung 44: Möglichkeiten zur sofortigen Unterbrechung/Auswertung ..</i>	<i>347</i>

Abkürzungsverzeichnis

CB	Computer Based
CLT	Cognitive Load Theory
CTML	Cognitive Theory of Multimedia Learning
DBR	Design-Based Research
d. h.	das heißt
etc.	et cetera
HCI	Human Computer Interaction
ICT	Information- and Computer-Technologie
IKT	Informations- und Kommunikations-Technologie
Kap.	Kapitel
MW	Mittelwert
PDAs	Personal Digital Assistants
PP	Paper Pencil
S.	Seite
SD	Standardabweichung
TBA	Technology Based Assessment
u. a.	unter Anderem
UE	User Experience
UEQ	User Experience Questionnaire
vgl.	vergleiche
WCAG	Web Content Accessibility Guidelines
z. B.	zum Beispiel

Einleitung

1.1 Motivation und Zielsetzung der Arbeit

In Deutschland leben 7,5 Millionen funktionale Analphabeten und Analphabetinnen zwischen 16 und 64 Jahren. Das entspricht 14,5% der Deutsch sprechenden Bevölkerung (Grotlüschen & Riekmann, 2011, S. 2). Funktionale Analphabeten und Analphabetinnen sind Personen, die zwar einzelne Buchstaben und Wörter lesen und schreiben können, nicht jedoch zusammenhängende Texte. Die betroffenen Personen sind „aufgrund ihrer begrenzten schriftsprachlichen Kompetenzen nicht in der Lage, am gesellschaftlichen Leben in angemessener Form teilzuhaben“ (Grotlüschen & Riekmann, 2011, S. 2). Auch die aktuelle PIAAC-Studie kommt zu ähnlichen Ergebnissen: 17,5% der deutschen Bevölkerung zwischen 16 bis 65 Jahren befinden sich demnach unter bzw. auf der niedrigsten Kompetenzstufe I (Zabal u. a., 2013, S. 42), im Bereich der alltagsmathematischen Kompetenz 18,4% unter bzw. auf Stufe I (Zabal u. a., 2013, S. 55). Deutschland bewegt sich damit im OECD-Durchschnitt.

Es besteht somit ein enormer Alphabetisierungsbedarf. Demgegenüber steht die Alphabetisierungspraxis allerdings vor diversen Herausforderungen. Aus der Perspektive der Betroffenen existieren vielfältige Hemmnisse, an einem Alphabetisierungskurs teilzunehmen (in Anlehnung an Steuten, 2013, S. 9; vgl. auch Wolf & Koppel, 2014):

1. Insbesondere außerhalb städtischer Ballungszentren fehlt es an ausreichenden Angeboten vor Ort (so bietet nur ungefähr jede dritte Volkshochschule Alphabetisierungskurse an) (Rosenblatt & Lehmann, 2013b, S. 57);
2. Alphabetisierungsangebote haben einen geringen Bekanntheitsgrad;
3. Betroffene meiden die Offenlegung („coming out“) empfundener und/oder tatsächlicher Schwächen (z. B. aufgrund von Schamgefühlen);
4. der Besuch von Alphabetisierungsanbietern wird als ähnlich unangenehm wie der Besuch von Behörden oder Bildungseinrichtungen empfunden;
5. Betroffene empfinden subjektiv keinen Bedarf an einer Literalitätsförderung.

Haben Betroffene den Weg in Alphabetisierungskurse gefunden und nehmen daran teil, existieren auf der Anbieterseite von Alphabetisierungsmaßnahmen weitere Herausforderungen, mit denen insbesondere Kursleitende konfrontiert sind:

1. Standardisierte Curricula existieren erst seit Mitte des Jahres 2014 (Ossner, 2014; Rosebrock, Gausche, Haase & Zimper, 2014 herausgegeben vom Deutschen Volkshochschulverband e.V.);
2. es fehlen einheitliche Lehr-/Lernmethoden (Rosenblatt & Lehmann, 2013b, S. 57);
3. die Vorstellung über die Abschlussziele sind nicht klar definiert (Rosenblatt & Lehmann, 2013b, S. 57);
4. erst seit wenigen Jahren wird die Forschung über und Entwicklung von erwachsenengerechtem Material praktiziert (Heinemann, 2011, S. 92);
5. der Einsatz diagnostischer Verfahren ist umstritten¹ (Bonna & Nienkämper, 2011);
6. Testsituationen stoßen bei den Teilnehmenden vielfach auf Abwehrreaktionen (vgl. z. B. Schladebach, 2007).

Schließlich verweist die Studie Alphapanel (von Rosenblatt & Bilger, 2011) auf kritische Aspekte hinsichtlich der Kosten und des Ertrags von Alphabetisierungsmaßnahmen (vgl. auch Rosenblatt & Lehmann, 2013a, 2013b):

1. Alphabetisierungskurse können aufgrund des hohen individuellen Betreuungsbedarfs nur in kleinen Gruppen mit qualifizierten Alphabetisierungskräften durchgeführt werden und sind somit personalintensiv und teuer.
2. Da funktionale Analphabeten und Analphabetinnen Lesen und Schreiben in unterschiedlichem Ausmaß gelernt (und auch wieder verlernt) haben, sind die Alphabetisierungskräfte mit einem Patchwork an Kompetenzen konfrontiert und bedürfen daher individueller Fördermaßnahmen. Die dazu notwendige individuelle Diagnostik des Lernstandes ist allerdings zeitaufwändig. Die Testanweisungen sind meist auf einem höheren schriftsprachlichen Niveau als die eigentlichen Testaufgaben. Zudem ist die manuelle Auswertung zeitintensiv. Daher ist eine individuelle Diagnostik nur in direkter 1:1-Betreuung durchführbar.
3. Bisher führte die Teilnahme an Alphabetisierungskursen nur zu einem begrenzten Lernerfolg: Auch wenn es Lernfortschritte gibt, reichen diese meist nicht aus, um bei der Mehrzahl der Teilnehmenden ein schriftsprachliches Kompetenzniveau oberhalb des funktionalen Analphabetismus zu erreichen.

Um diesen Herausforderungen sowohl auf der Seite der Betroffenen (der funktionalen Analphabeten und Analphabetinnen) als auch der Anbieter zu begegnen, kann eine online zugängliche computerbasierte Diagnostik einen Beitrag zur

¹ Vermutet wird von den Autorinnen der hier zitierten Akzeptanzstudie, dass in der Alphabetisierung eine förderdiagnostische Praxis im Vergleich zu einer selektiven Diagnostik bevorzugt wird (Bonna & Nienkämper, 2011, S. 46).

Alphabetisierung leisten: Einerseits können Betroffene das Instrument selbstständig und anonym nutzen und müssen sich somit nicht einer Präsenzsituation stellen. Andererseits unterstützt sie Alphabetisierungskräfte, indem eine individuelle Rückmeldung ohne eine 1:1-Betreuung möglich ist. In dem Projekt lea. – Literalitätsentwicklung für Arbeitskräfte² – wurde daher neben einer papierbasierten Version eine Online-Testumgebung – otu.lea³ – für funktionale Analphabeten und Analphabetinnen entwickelt. otu.lea bietet eine Kompetenzdiagnose in den Bereichen Lesen, Schreiben, mathematisches Grundwissen sowie Sprachempfinden. Die Diagnostik deckt den Bereich der sog. Lower Rungs – unterhalb bzw. auf der untersten Stufe der internationalen Vergleichsstudien PISA (Programme for International Student Assessment) und IALS (International Adult Literacy Survey) – ab. Aus der Perspektive der Nachfrageseite bzw. der Betroffenen kann otu.lea den Herausforderungen durch

- den Zugang über das Internet und somit unabhängig von einem flächendeckenden Angebot, und
- die Möglichkeit der anonymen Nutzung und somit ohne eine notwendige Offenlegung der Literalitätskompetenzen begeben.

Die Anbieterseite wird von otu.lea durch

- die Anknüpfung an die aktuellen Curricula,
- eine differenzierte Rückmeldung über die individuellen Kompetenzen (bzw. über das Kompetenzpatchworks),
- die Möglichkeit, anhand der Kompetenzstufen und -definitionen klare Abschlussziele zu definieren und zu überprüfen,
- die Einbindung der Lebens- und Arbeitswelt von Erwachsenen,
- eine stärkenorientierte und im Idealfall lernmotivationsfördernde Rückmeldung,
- die Möglichkeit des selbstständigen Gebrauchs, so dass ein ressourcenschonender Einsatz in Alphabetisierungskursen möglich ist,

unterstützt.

Die Aufbereitung für den selbstständigen Gebrauch stellte die Entwickler und Entwicklerinnen allerdings vor diverse Herausforderungen. Die Inhalte waren zwar durch die bereits entwickelte papierbasierte Version vorgegeben, doch wäre eine direkte 1:1-Übertragung von der papierbasierten in eine elektronische Version den Anforderungen einer inhaltlich validen gebrauchstauglichen Version nicht gerecht geworden. Funktionale Analphabeten und Analphabetinnen haben neben ihren niedrigen schriftsprachlichen Kompetenzen zudem eine tendenziell niedrige Computerkompetenz (Niesyto, 2009, S. 6; Zabal u. a., 2013, S. 68). Wie

² <http://blogs.epb.uni-hamburg.de/lea/> (zuletzt geprüft am 03.11. 2014)

³ www.otulea.uni-bremen.de (zuletzt geprüft am 03.11. 2014)

muss demnach eine computerbasierte Diagnostik für funktionale Analphabeten und Analphabetinnen gestaltet sein, damit die Computerkompetenz keinen Einfluss auf das Testergebnis hat? Wenn die Online-Diagnostik anonym und selbstständig nutzbar sein soll, muss sie leicht erschließbar und ohne außenstehende Erläuterungen verständlich sein. Instruktionen und Aufgabeninhalte können von anderen Personen nicht vorgelesen oder erläutert werden. Daher ist bei otu.lea auf eine ausgeprägte Usability (engl. für Gebrauchstauglichkeit) zu achten. Die Usability bezeichnet dabei „das Ausmaß, in dem ein Produkt durch bestimmte Benutzer und Benutzerinnen in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und mit Zufriedenheit zu erreichen.“ (DIN EN ISO, 2011, S. 38). Das Konzept der Usability ist somit relational: Die Ausprägung der Usability ist vom jeweiligen Nutzungskontext abhängig. Hierzu zählen neben den technischen Rahmenbedingungen und dem Einsatzbereich auch die Nutzer und Nutzerinnen selbst. Bisher existieren allerdings wenig Hinweise darauf, wie die Usability speziell für funktionale Analphabeten und Analphabetinnen gewährleistet werden kann. Auch existieren bisher keine Studien mit dieser Zielgruppe über die Nutzung von Softwareprodukten. Die Web Content Accessibility Guidelines (WCAG) (deutsch: Richtlinien für barrierefreie Webinhalte) haben zum Ziel, Gestaltungsrichtlinien zu entwickeln, um Webinhalte für jegliche Nutzer- und Nutzerinnengruppen zugänglich zu machen. Sie fokussieren dabei insbesondere Personen mit sensorischen und motorischen Einschränkungen (Caldwell, Cooper, Reid & Vanderheiden, 2008). Da funktionale Analphabeten und Analphabetinnen aber weder durch das Fehlen sensorischer noch motorische Kompetenzen charakterisiert sind, ist nicht nur zu bezweifeln, dass die WCAG auf funktionale Analphabeten und Analphabetinnen übertragbar sind, sondern lässt sich vielmehr an der Akzeptanz solcher, nicht an den spezifischen Anforderungen der Nutzer und Nutzerinnen angepassten Software, zweifeln. Für den englischsprachigen Raum hat das National Research Development Centre for adult literacy and numeracy (NRDC) in England einen ausführlichen Report über bestehende literalitätsbezogene computerbasierte Diagnoseprogramme in England vorgelegt. Das NRDC kommt zu dem Ergebnis, dass bestehende computerbasierte Diagnoseinstrumente hinsichtlich ihrer Usability unzureichend sind. Insbesondere gilt dies für Instrumente, die primär von Personen mit einer niedrigen *Information and Communication Technology (ICT) -Literacy* genutzt werden (Brooks, Heath & Pollard, 2005, S. 21).

An diese Herausforderungen anknüpfend ist das Ziel des vorliegenden Forschungsvorhabens die Entwicklung und Evaluation einer erwachsenengerechten computerbasierten Online-Diagnostik für den Grundbildungsbereich. Unter Berücksichtigung des aktuellen Forschungsstands ist die Beantwortung der Forschungsfrage „Wie muss eine Online-Diagnostik für funktionale Analphabeten

und Analphabetinnen für den selbstständigen Gebrauch gestaltet sein?“ über den Zugang aktueller Erkenntnisse über computerbasierte Diagnostik, zielgruppenspezifische Charakteristika zur Computerkompetenz und -nutzung, kognitionspsychologische Aspekte sowie allgemeinen über Usability zu wählen. Vor dem Hintergrund dieser Erkenntnisse wurde das Konzept für otu.lea erstellt, iterativ evaluiert und weiterentwickelt. Für das Forschungsvorhaben wurden neben laufenden Evaluationen drei umfangreiche Evaluationsstudien durchgeführt. Dessen Ergebnisse wurden analysiert und interpretiert, um sie in den jeweiligen anschließenden Modifikationsphasen zu berücksichtigen und otu.lea weiterzuentwickeln. Die Studien fanden in den Jahren 2010, 2012 und 2013 statt. Dabei kamen sowohl qualitative (Interviews, Videos) als auch quantitative (Data-Logging für die Messung von Performanz und Bearbeitungszeit, Befragungen zur Computererfahrung und -kompetenz sowie Bewertungsbögen für otu.lea) Erhebungsmethoden zum Einsatz.

1.2 Aufbau der Arbeit

Als Forschungsrahmen für die Umsetzung des Forschungsvorhabens dient der Design-Based Research (DBR)-Ansatz. Dieser hat zum Ziel, durch die systematische Gestaltung, Durchführung, Evaluation und Re-Design innovative Lösungen für Probleme in der Bildungspraxis zu entwickeln (vgl. z. B. Edelson, 2002). Charakteristisch ist dabei eine iterative Vorgehensweise, in der sowohl Forschende als auch Praktizierende in den Entwicklungsprozess einbezogen werden (vgl. z. B. T. Anderson & Shattuck, 2012). Der DBR-Prozess sieht ein Vorgehen in vier Phasen vor (Reeves, 2006): Problemanalyse, Konzeption, Konzeptevaluation sowie Reflexion. Dabei stellt die Phase der Konzeptevaluation einen Zyklus aus vier Schritten dar: Gestaltung, Durchführung, Analyse und Re-Design. Entsprechend des DBR-Ansatzes gliedert sich die vorliegende Arbeit in vier Phasen.

Phase I – Problemanalyse: In der ersten Phase wird der aktuelle Forschungsstand dargestellt und die theoretische Basis für die anschließende empirische Untersuchung geschaffen (Kapitel (Kap. 2). Beginnend wird das Thema Diagnostik aus mehreren Perspektiven aufgegriffen: Nach einer kurzen Abhandlung der Pädagogischen Diagnostik und der Förderdiagnostik (Kap. 2.1) wird umfassend auf die computerbasierte Diagnostik eingegangen (Kap. 2.2). Es werden sowohl Begrifflichkeiten erläutert als auch aktuelle Forschungsentwicklungen computerbasierter Diagnostik und deren Gestaltungsmöglichkeiten aufgegriffen. Um die zielgruppenspezifischen Herausforderungen in der Entwicklung und Evaluation einbeziehen zu können, werden im Kap. 2.3 die Themen Literalität und Alphabetisierung erörtert. Für die Gestaltung einer gebrauchstauglichen Testumgebung sind zudem ein kognitionspsychologischer Zugang gewählt. Hierfür werden aktuelle Erkenntnisse der *Cognitive Load Theory* (CLT) und der

Cognitive Theory of Multimedia Learning (CTML) (Kap. 2.4) berücksichtigt. Anschließend wird in Kap. 2.5 auf die Äquivalenzproblematik eingegangen, indem die Problematik der Übertragung eines papierbasierten in ein elektronisches Instruments erörtert wird. Die bis dahin gelegten theoretischen Grundsteine werden schließlich in der Bearbeitung des Themas Usability berücksichtigt (Kap.2.6). Für die Auseinandersetzung mit der Frage nach einer gebrauchstauglichen Gestaltung für funktionale Analphabeten und Analphabetinnen werden die Begrifflichkeiten, die Anforderungen an eine ausgeprägte Usability sowie deren Evaluationsmöglichkeiten erörtert.

Phase II – Konzeption: In der zweiten Phase werden zum Einen die Konzeption der empirischen Untersuchung und zum anderen die Konzeption des Untersuchungsgegenstands – die Online-Testumgebung – erläutert.

Für die empirische Untersuchung stellt der Design-Based Research-Ansatz den Forschungsrahmen und damit die methodologische Grundlage der empirischen Untersuchung dar (Kap. 3). Hierfür werden die theoretischen Grundlagen, typische Anwendungskontexte sowie kritische Aspekte dargestellt. Anschließend wird die Konzeption der Online-Testumgebung vorgestellt, indem der Projektkontext (Kap. 4.1) sowie die Entwicklung der Online-Testumgebung otu.lea (Kap. 4.2) dargestellt werden.

Phase III – Konzeptevaluation: Im Rahmen der Konzeptevaluation wird der Forschungsrahmen von DBR auf das Forschungsvorhaben übertragen (Kap. 5). Die Konzeptevaluation besteht aus drei Design-Zyklen: Einem Paper Prototyping-Experiment im Jahr 2010 (Kap. 6), einer Usability-Studie im Jahr 2012 (Kap. 7) sowie einer abschließenden ausführlichen Evaluationsstudie im Jahr 2013 (Kap. 8). Die Kapitel zu den Evaluationszyklen (Kap. 6 bis Kap. 8) sind jeweils in die Evaluationsschritte Gestaltung, Durchführung, Analyse und Re-Design gegliedert. Der Fokus der empirischen Erhebungen liegt dabei auf der dritten Studie. In dieser wurde das Online-Diagnose-Instrument unter Einsatz verschiedener Erhebungsinstrumente mit 50 Probanden und Probandinnen evaluiert und modifiziert. Im Anschluss an die Ergebnisdarstellung und -interpretation werden Schlussfolgerungen sowohl auf der theoretischen Ebene für die Gestaltung und des Einsatzes einer Online-Diagnostik für den Grundbildungsbereich als auch auf der praktischen Ebene für die Entwicklung der Online-Diagnostik otu.lea gezogen.

Phase IV – Reflexion: In der vierten Phase erfolgt schließlich die Reflexion des Forschungsprozesses. Hierfür werden der Forschungsrahmen des DBR (Kap. 9.1) sowie die Erhebungsmethoden und -instrumente (Kap. 9.2) einer kritischen Reflexion unterzogen. Im neunten Kap. folgen eine abschließende zusammenfassende Diskussion der Ergebnisse, ein Ausblick und das Aufzeigen von Forschungsdesiderata.

Phase I – Problemanalyse

2 Theoretische Herleitung

Die theoretische Basis für die Beantwortung der Forschungsfrage stellen aktuelle Erkenntnisse aus der pädagogischen Diagnostik, der Kognitionspsychologie, und insbesondere der *Cognitive Load Theory* (CLT), sowie der Usability-Forschung dar. In den folgenden Kapiteln wird die Forschungsfrage aus den eben genannten theoretischen Zugängen hergeleitet.

2.1 Diagnostische Grundlagen

Die Online-Diagnostik *otulea* soll sowohl für genuin diagnostische Tätigkeiten als auch für Tätigkeiten im Kontext förderdiagnostischer Absichten entwickelt werden. In den folgenden zwei Kapiteln werden daher grundlegende Erkenntnisse zur Pädagogischen Diagnostik sowie Prinzipien der Förderdiagnostik erläutert.

2.1.1 Pädagogische Diagnostik

Die Pädagogische Diagnostik umfasst Tätigkeiten, die bei den Lernenden die Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse ermitteln, Lernprozesse analysieren sowie Lernergebnisse feststellen (Ingenkamp & Lissmann, 2008, S. 13). Die Ergebnisse der Pädagogischen Diagnostik dienen der Zuweisung zu Lerngruppen oder individuellen Förderprogrammen, um beispielsweise durch weniger heterogene Gruppen ein "effizienteres" Lernen und Fördern zu ermöglichen. Übergeordnetes Ziel der Pädagogischen Diagnostik ist somit die Optimierung individuellen Lernens. Zudem dient sie der Steuerung des Bildungsnachwuchses, z. B. durch die differenzierte Abbildung des Kompetenzstands und anschließender Empfehlungen für den weiteren Bildungsweg oder der Erteilung von Qualifikationen, beispielsweise mittels der Bestätigung eines bestimmten Kompetenzniveaus.

Zunehmend wurden und werden unter den Funktionen der pädagogischen Diagnostik auch die Förderung des Lernens, die Verbesserung des Unterrichts, die Schüler-/Schülerinnenberatung und die Platzierung im Bildungssystem genannt (Lindquist, 1950; zitiert nach Ingenkamp & Lissmann, 2008, S. 22). Weiterhin werden als Ziele des Einsatzes von diagnostischen Verfahren die Diagnose von Stärken und Schwächen des Lernens, die Planung und Verbesserung des Unterrichts sowie die Evaluation von Leistung genannt (Phye, 1996; zitiert nach Ingenkamp & Lissmann, 2008). Die Pädagogische Diagnostik kann also auch einen Bestandteil einer Evaluation darstellen, doch bezieht sich der Begriff Evaluation primär auf die Bewertung von Interventionen (Wottawa, 2006, S. 650),

beispielsweise Kursen oder Kursentwicklungen oder auch auf den Prozess, wie sich die Bewertungen und das Bewertungsschema entwickelt haben (Taras, 2005, S. 467) und weniger auf die Bewertung personenbezogener Merkmale.

Bei der Pädagogischen Diagnostik handelt es sich in ihrer Konsequenz meist um selektive Tätigkeiten und somit um selektive Diagnostik. In der selektiven Diagnostik, als spezielle Ausprägung der pädagogischen Diagnostik, wird eine soziale oder sachliche Bezugsnorm herangezogen, um entweder zu prüfen, wie sich die Performanz einer Person im Vergleich zu einer anderen Person bzw. Gruppe oder zu einem fachlich definierten Anspruch verhält. Zudem wird die selektive Diagnostik von einer außen stehenden Instanz (z. B. Lehrperson) durchgeführt, bei der die Ergebnisse verbleiben (Schule, Arbeitgeber, Arbeitsagentur etc.). Als Konsequenz der selektiven Diagnostik wird – wie der Name ausdrückt – selektiert, d. h. es wird eine soziale oder sachliche und keine individuelle Bezugsnorm für die Leistungsbewertung herangezogen und eine Person wird aufgrund ihrer Performanz einer bestimmten Gruppe oder auch Institution zugeordnet. Die/Der Lernende steht dabei nicht im Fokus, was wiederum einen negativen Effekt auf die Lernmotivation haben kann (vgl. z. B. Rheinberg & Krug, 2005). Historisch betrachtet gewann die pädagogische Diagnostik zur Erteilung von Qualifikationen (wie z. B. die Erteilung von Zugangsberechtigung zur Ausbildung/zum Studium) in den letzten zwei Jahrhunderten zunehmend an Bedeutung⁴. Mit der zunehmenden Demokratisierung und der Ausschaltung von Geburtsrechten und Ämterkauf wurden individuelle Lernerfolge zunehmend wichtiger. Allerdings verlief eine Diagnostik bis dahin eher intuitiv und nicht mit Hilfe wissenschaftlicher Methoden. Ein vielfältiges Instrumentarium möglichst objektiver Verfahren ist in den vergangenen hundert Jahren entstanden (Ingenkamp & Lissmann, 2008, S. 22), mit Hilfe derer (unter Berücksichtigung der Gütekriterien) Beobachtungs- und Befragungsergebnisse ausgewertet werden, um die aktuelle und/oder zukünftige Performanz zu beschreiben. Ein Beispiel, das bis in die Gegenwart hineinreicht, ist der Einsatz von Schulleistungstest. Von der Pädagogischen Diagnostik abzugrenzen ist die pädagogische Forschung. In der pädagogischen Forschung wird diagnostiziert, um allgemeine Erkenntnisse über beispielsweise die Angemessenheit bestimmter didaktischer Vorgehensweisen oder den Mediengebrauch zu gewinnen. Auch wenn hier diagnostische Methoden in der Pädagogik genutzt werden, sind diese Erhebungen nicht der pädagogischen Diagnostik zuzuordnen. Die pädagogische Diagnostik zeichnet sich somit – im Unterschied zur pädagogischen Forschung – durch den Bezug auf aktuelle Maßnahmen bei einzelnen Lernenden oder einer Gruppe von Lernenden aus (Ingenkamp & Lissmann, 2008, S. 19). Im internati-

⁴ Wobei anzumerken ist, dass auch bereits zu Beginn des 7. Jahrhunderts in China Beamtenprüfungen durchgeführt wurden (Glöckner, 2013, S. 194).

onalen Raum ist für den Begriff der Diagnostik im Allgemeinen der Begriff *Assessment* (engl. für *Bewertung*) gebräuchlich. Der Begriff *Diagnostic Assessment* ist am ehesten mit pädagogischer Diagnostik zu übersetzen. *Diagnostic Assessment* wird durchgeführt, um Vorwissen von Lernenden zu erheben und um zukünftiges Lernen zu optimieren (Crisp, 2007, S. 254). Um explizit auf einen pädagogischen Bezug hinzuweisen wird auch der Begriff *Educational Assessment* gewählt, jedoch wird von der Spezifizierung „*Educational*“ selten Gebrauch gemacht⁵. Weniger im deutschen, jedoch im englischen Sprachraum werden differenzierend, u.a. im pädagogischen Kontext, die Begriffe *Large-Scale Assessment* und *Summative Assessment* genutzt. *Large-Scale Assessment* sind Schulleistungsuntersuchungen mit einer großen Anzahl von Teilnehmenden.

„Large-scale educational assessment consists of those tests administered to sizable numbers of people for such purposes as placement, course credit, graduation, educational admissions, and school accountability. It includes group-administered, standardizes tests“ (Bennett, 1998, S. 7).

Large-Scale Assessments zeichnen sich demnach durch ähnliche Ziele wie die Pädagogische Diagnostik aus, allerdings mit der Bedingung große Stichproben einzusetzen. Charakteristisch ist dabei die Durchführung von vorwiegend politischen Entscheidungsträgern und Bildungspolitikern bzw. führenden Bildungsinstitutionen, um entweder Bildungsprogramme zu evaluieren oder um Lernziele bei Schülern und Schülerinnen festzustellen (Pellegrino, Chudowsky & Glaser, 2001). Die Datenhoheit liegt dabei bei der forschenden Instanz. Ein bekanntes Beispiel ist die internationale Vergleichsstudie PISA (*Programme for International Student Assessment*).

Beim *Summative Assessment* handelt es sich in seiner Konsequenz meist um eine selektive Diagnostik.

„Summative assessment is intended to summarize student accomplishment by making judgment or determining a grade“ (Stödsberg, 2012, S. 595).

Summative Assessment bedeutet, Hinweise und Ergebnisse hinsichtlich eines vorgegebenen Standards auszuwerten. Dieser Punkt hat in Bezug auf *Assessment* den Charakter der Endgültigkeit. Der Prozesscharakter wird beim *Summative*

⁵ So findet sich z. B. im Glossar des E-Assessment Handbooks (Crisp, 2007) zwar eine Definition des Begriffs "Diagnostic Assessment", jedoch nicht für "Educational Assessment".

Assessment nicht berücksichtigt (Taras, 2005, S. 468). Sowohl das *Large-Scale* als auch das *Summative Assessment* haben eine soziale Bezugsnorm (im Gegensatz zum *Formative Assessment*, welches auf eine individuelle Bezugsnorm referenziert und nicht die einmalige Bewertung einer Person zum Zwecke der Selektion zum Ziel hat (vgl. nachfolgendes Kap. 2.1.2) und erheben den Anspruch der Statuserhebung.

Der deutsche Begriff Pädagogische Diagnostik scheint somit eine eindeutige Domänenspezifität (nämlich Pädagogik) vorzuweisen. Im Englischen werden mit dem Begriff *Assessment* nicht nur pädagogische Fragestellungen assoziiert. Die Differenzierungen in *Summative* und *Large-Scale Assessment* weisen hingegen eine Spezifität auf, zu denen sich im Deutschen keine äquivalenten Begriffe finden lassen. Die englischen Begriffe wurden in den deutschen Sprachraum übernommen, wobei der englische Begriff *summative* in „summativ“ übersetzt wurde, für den englischen Begriff *Large-Scale* besteht allerdings keine äquivalente deutsche Übersetzung. Dieser Abgrenzungsansatz zwischen deutschen und englischen Begriffen sowie deren Ausprägungen macht deutlich, dass Begriffe aus dem Englischen verwendet werden, deren Verwendung hingegen im Deutschen nicht immer äquivalent stattfindet. Um ein einheitliches und präzises Verständnis der Begrifflichkeiten zu ermöglichen, werden in den folgenden Kapiteln vorzugsweise die deutschen Begriffe verwendet. In einigen Ausnahmefällen wird auf englische Begriffe zurückgegriffen, da es sich entweder um im deutschen Sprachraum etablierte Anglizismen handelt und/oder der englische Begriff ein präziseres Verständnis ermöglicht.

Zusammenfassend liefert die pädagogische Diagnostik zwar Ergebnisse zum Leistungsstand einer Person, allerdings nicht (bzw. nur sehr bedingt) hinsichtlich abzuleitender individueller Fördermaßnahmen. Faktoren, die außerhalb der Prüfungsleistung liegen, wie z. B. Lernhemmungen, -behinderungen, das soziale Umfeld oder fehlende Passungen zwischen Unterrichtsmethode und Test, werden ebenfalls nicht berücksichtigt. Diese Abgrenzung der Funktionen und Ziele pädagogischer Diagnostik macht deutlich, dass der Bereich der pädagogischen Diagnostik auch das Thema Förderung umfasst, allerdings ohne explizit der Förderdiagnostik gerecht zu werden (vgl. folgendes Kap. 2.1.2). Diese fehlenden Faktoren werden in der Pädagogischen Förderdiagnostik einbezogen (Engel, 2008, S. 33). Die Pädagogische Diagnostik ist damit keineswegs als Gegenpool zur Förderdiagnostik zu verstehen, vielmehr stellt Pädagogische Diagnostik eine Erweiterung zur pädagogischen und speziell zur selektiven Diagnostik sowie eine Spezifizierung (u. a. in der Darstellung der individuellen Performanz) dar.

2.1.2 Förderdiagnostik

Die pädagogische Förderdiagnostik geht über reine diagnostische Tätigkeiten hinaus, um nach Ansatzpunkten für eine Förderung sowie nach veränderbaren Bedingungen in der Lernsituation zu suchen (Dupuis & Kerkhoff, 1992, S. 212). Pädagogische Förderung besteht in der Bereitstellung und Durchführung individuell zugeschnittener Angebote, wenn die (schulischen) Standardangebote nicht ausreichend für die Entwicklung der Person sind (Kretschmann, 2006, S. 140). Die in der pädagogischen Diagnostik gewonnenen Informationen werden genutzt, um individuelle Fördermaßnahmen abzuleiten (Breitenbach, 2007, S. 40), wobei die Daten bei der diagnostizierenden Instanz verbleiben. Sechs Prinzipien sind in der Förderdiagnostik vorherrschend (in Anlehnung an Rittmeyer, 2005; Schönrade & Pütz, 2004; vgl. Wolf, Koppel & Schwedes, 2011):

1. Individualität
2. Prozessorientierung
3. Wechselverhältnis von Diagnose und Intervention
4. Einbettung in das reale Umfeld der Teilnehmenden
5. Berücksichtigung von Stärken und Schwächen
6. Anwendung von Beobachtungsverfahren und Fehleranalysen

Im Folgenden werden die Prinzipien näher erläutert:

1. Individualität: Wie bereits vorangegangen angedeutet ist die Förderdiagnostik am individuellen Lernenden orientiert. Ziel ist, Lernschwierigkeiten und deren Entstehung zu ergünden. Bei Bedarf werden Veränderungen von Lernprozessen initiiert. Der Förderbedarf wird deskriptiv erfasst und es werden Hypothesen über mögliche Ursachen für Lernschwierigkeiten gebildet, um mögliche Interventionsstrategien zu entwickeln (Engel, 2008, S. 34).

2. Prozessorientierung: Im Kontext förderdiagnostischen Handelns wird davon ausgegangen, dass Entwicklungs- und Lernfähigkeit des Lernenden dynamisch veränderbar sind. Daher ist die Förderdiagnostik immer über einen längeren Zeitraum durchzuführen. In diesem werden mehrmals Lernstandserhebungen vorgenommen und Fördermaßnahmen abgeleitet (Petermann & Petermann, 2006, S. 2). Folglich werden die Begriffe Lernprozessdiagnostik und Lernprozessanalyse verwendet (Belusa & Eberwein, 1997). Wichtiges Kriterium ist zudem die Einbindung laufender Feedbackprozesse, die eine Rückmeldung über die Diskrepanz zwischen der aktuellen Performanz und der Zielvorgabe geben. Dies hat u.a. die Förderung des Lernerfolgs als auch der Lernmotivation zur Folge (Nicol & Milligan, 2006; Rheinberg & Krug, 2005).

3. Wechselverhältnis von Diagnose und Intervention: Diagnose und Intervention stehen in der Förderdiagnostik in einem Wechselverhältnis. Für die Ableitung von Fördermaßnahmen ist die Diagnose notwendig, um an den individuellen Leistungsstand anknüpfen zu können. Nach Einsatz der Fördermaßnahmen

wird der Erfolg der Interventionsmaßnahmen überprüft, um anhand der evtl. gesteigerten Performanz neue Fördermaßnahmen zu gestalten oder bei gleichbleibender Performanz die Fördermaßnahmen zu modifizieren. Bei mehrfachen Sequenzen von Förderung und Intervention entsteht so ein förderdiagnostischer Regelkreis (Rittmeyer, 2005, S. 18 in Anlehnung an Schönrade & Pütz, 2004).

4. Einbettung in das reale Umfeld der Teilnehmenden: Viele relevante Informationen lassen sich aus Beobachtungen in Alltagssituationen gewinnen. Indem die Informationsgewinnung im Alltag stattfindet, wird das Umfeld der Teilnehmenden berücksichtigt und in die Förderdiagnostik einbezogen, wie beispielsweise die zeitlichen Ressourcen oder die Einstellung zu Fördermaßnahmen. Förderdiagnostik ist somit nicht nur Lernprozess-, sondern auch Situationsdiagnostik (Breitenbach, 2007, S. 25; Engel, 2008, S. 34; Rittmeyer, 2005, S. 20).

5. Berücksichtigung von Stärken und Schwächen: Die Förderdiagnostik diagnostiziert neben dem Förderbedarf auch die Stärken des Lernenden, beispielsweise dessen Kompetenzen und Fähigkeiten (Eggert, 1997) und ist damit nicht nur defizit- sondern auch bzw. insbesondere stärkenorientiert.

6. Anwendung von Beobachtungsverfahren und Fehleranalysen: Instrumente für die Ermittlung des Förderbedarfs sind Beobachtungsverfahren und Fehleranalysen (Rittmeyer, 2005, S. 29). Beobachtungsverfahren haben das Ziel, große Einheiten des Verhaltens und Erlebens zu erfassen und finden meist im natürlichen Lebensumfeld der beobachteten Personen statt (Bortz & Döring, 2006, S. 322). Bei den Fehleranalysen wird die Identifikation von systematischen Fehlern fokussiert, da diese sowohl Rückschlüsse auf den Lernstand als auch Hinweise auf konkrete Ansatzpunkte zur Förderung liefern.

Im englischen Sprachraum ist für Förderdiagnostik der Begriff *Formative Assessment* gebräuchlich, wobei in der englischen Begriffsbestimmung eine weitere Dimension berücksichtigt wird: Die Zusammenarbeit zwischen Lehrenden und Lernenden sowie zwischen den Lernenden untereinander (McManu, 2008). Für eine gelingendes *Formative Assessment* soll eine partnerschaftliche Lernatmosphäre sowohl zwischen Lernenden und Lehrenden bestehen als auch zwischen den Lernenden selbst, um ihre metakognitiven Lernstrategien und somit die Reflexionsfähigkeit zu verbessern.

2.1.3 Zusammenfassender Vergleich und Akzeptanz diagnostischer Verfahren

Die Förderdiagnostik lässt sich von der selektiven Diagnostik anhand von sechs Dimensionen abgrenzen (in Anlehnung an Dluzak, Heinemann & Grotluschen, 2009).

Dimension	Selektive Diagnostik	Förderdiagnostik
Bezugsnorm	sozial	individuell
Datenhoheit	diagnostizierende Instanz	diagnostizierende Person
Konsequenz	Selektion	Anpassung der Lernangebote
Instanz	Fremdbeurteilung	Fremd-, Peer- oder Selbstbeurteilung
Perspektive	Statuserhebung	Status- oder Prozesserhebung
Zeitpunkt in Relation zum Bildungsangebot	vorlaufend und/oder nachlaufend	vorlaufend und/oder mitlaufend

Tabelle 1: Dimensionsausprägung selektive Diagnostik und Förderdiagnostik (in Anlehnung an Dluzak u. a., 2009, S. 34)

Die Förderdiagnostik findet in vielen Bereichen Anwendung (Schule, Weiterbildungseinrichtungen) und fokussiert unterschiedliche Altersgruppen (Schüler/Schülerinnen, Erwachsene). Bezüglich der Anwendung diagnostischer Verfahren waren viele Jahre die Überlegungen vorherrschend, Erwachsene im Grundbildungsbereich würden sich einer Diagnostik (auch wenn sie zur weiteren Förderung verwendet werden soll), aufgrund der Befürchtung weiterer Defiziterfahrungen, verwehren (Füssenich, 2004). Diese Annahmen werden durch aktuelle Erfahrungsberichte (Brigitte, 2004) und Studien (Nienkämper & Bonna, 2010; Schladebach, 2007) widerlegt. So wird in der Akzeptanzstudie zur Akzeptanz von Diagnostik in der Alphabetisierung (2008-2011) berichtet, dass ca. 75% der befragten Kursleiter und Kursleiterinnen angeben, Diagnosematerialien einzusetzen, deren Anwendung nicht zu lange dauert; 75% setzen Diagnosematerialien ein, bei denen das Verhältnis zwischen Aufwand und Ergebnis ausgeglichen ist (Bonna & Nienkämper, 2011, S. 46).

Dennoch ist davon auszugehen, dass gerade Erwachsene im Grundbildungsbereich aufgrund ihrer niedrigen Lese- und Schreibfähigkeiten schamhaft sind (Döbert, Hubertus & Nickel, 2000; Egloff, 1997; Füssenich, 2004; Schladebach, 2007). Betroffene berichten von einem Schamgefühl, sich einer Testsituation, beispielsweise in einer Institution wie die VHS, zu stellen. Dies hat eine niedrige Teilnehmer-/Teilnehmerinnenquote an Maßnahmen zu Alphabetisierung zur Folge: Obwohl die Anzahl funktionaler Analphabeten und Analphabetinnen in Deutschland 7,5 Millionen beträgt (Grotluschen & Riekman, 2011) nehmen nur ca. 30.000 Personen an Alphabetisierungskursen teil (Huntemann & Reichart, 2011, S. 32). Daher ist es wünschenswert, den Einstieg in die Teilnahme an Maßnahmen zur Alphabetisierung zu erleichtern (Wolf u. a., 2011, S. 127). Hilfreich kann hierbei die computerbasierte Förderdiagnostik sein, die einen flexiblen und anonymen Einsatz ermöglicht.

Zusammengefasst ist das Ziel der Diagnostik die Bewertung hinsichtlich Standards, Ziele und Kriterien. Erweiternd beinhaltet Förderdiagnostik ein Feedback, welches die Lücke zwischen dem aktuellen Fähigkeits- oder Arbeitsstand und den zu erreichenden Standard aufzeigt. Zudem impliziert die Förderdiagnostik Hinweise darauf, wie diese Lücke zwischen Leistungsstand und Zielvorgabe geschlossen werden kann. Unterscheidungsmerkmal zwischen der Selektions- und der Förderdiagnostik ist nicht das zur Diagnostik eingesetzte Instrument, sondern *wie* es eingesetzt wird.

2.2 Computerbasierte Diagnostik

Diverse Diagnostikverfahren werden computergestützt angeboten. In dem folgenden Kap. findet zunächst eine begriffliche Einordnung statt (Kap. 2.2.1). Anschließend werden die Anreicherungsmöglichkeiten durch Multimedia beschrieben (Rich E-Assessment, Kap. 2.2.2), um in den darauf folgenden Kapiteln auf die Vor- und Nachteile computerbasierter Diagnostik (Kap. 2.2.3) sowie auf die Potenziale im förderdiagnostischen Kontext (Kap. 2.2.4) eingehen zu können. Nachdem aktuelle Beispiele vorgestellt wurden (Kap. 2.2.5), werden schließlich innovative Itemformate der computerbasierten Diagnostik hinsichtlich möglicher Dimensionen erläutert (Kap. 2.2.6.1) und klassifiziert (Kap. 2.2.6.2).

2.2.1 Computerbasierte Diagnostik – Definition und Abgrenzung

Computerbasierte Diagnostik ist eine Subform des computerbasierten Assessments⁶ (engl. *Computer Based Assessment*) (vgl. z. B. Ruedel & Mandel, 2010). Jegliche computerbasierte Diagnostik ist also immer auch computerbasiertes Assessment. Um ein hinreichendes Verständnis bezüglich des Forschungsgegenstandes dieser Arbeit zu gewährleisten, wird im Folgenden zunächst der Begriff computerbasiertes Assessment erläutert, um daran anschließend den Begriff der computerbasierten Diagnostik zu erläutern.

Assessment bedeutet zunächst, mit Hilfe von statistisch erhobenen Daten Aussagen über den Wissenstand einer Person/Gruppe und über die Fähigkeiten einer Person/Gruppe treffen zu können. Die erhobenen Daten dienen der Bewertung von Kriterien hinsichtlich eines bestimmten Ziels (Crisp, 2007, S. 253). *Computerbasierte* Assessments bezeichnen spezifizierend den Einsatz von Computern zur Messung und Bewertung personaler Merkmale und beinhalten immer

⁶ Im Kontext computerbasierter Datenerhebung hat sich im deutschen Sprachraum der aus dem englischen Sprachraum stammende Begriff *Assessment* etabliert. Da in der Fachliteratur Begriffsausprägungen und –abgrenzungen anhand des Begriffs *Assessment* vorgenommen werden, wird auch in diesem Kapitel im Sinne eines präzisen und einheitlichen Verständnisses der Begriff *Assessment* verwendet.

die Nutzung von Informations- und Computertechnologie zur Erhebung der Daten und folglich der Bewertung von Kriterien. Sie umfassen den gesamten Prozess vom Aufgabendesign bis hin zur Datenspeicherung und -auswertung (Stöckberg, 2012, S. 591). Synonym verwendet werden oftmals die Begriffe elektronisches Assessment (engl. *E-Assessment*) oder technologiebasiertes Assessment (engl. *Technology Based Assessment*), wobei dies unter strenger Berücksichtigung der Bedeutungen der einzelnen Begrifflichkeiten nicht zutreffend ist. Technologiebasiertes oder elektronisches Assessment beschreibt den Einsatz jeglicher Informationstechnologie in der psychologischen und pädagogischen Diagnostik (Jurecka & Hartig, 2007, S. 37). Auch wenn in den meisten Fällen computerbasierte Assessments mit einem Desktop-Computer oder einem Laptop durchgeführt werden, muss dies nicht immer der Fall sein. Beispielsweise werden auch mobile Endgeräte wie Smartphones oder Personal Digital Assistants (PDAs) verwendet. Letztere sind z. B. von Vorteil, wenn Personen über einen längeren Zeitraum und in ihrem natürlichen Tagesablauf befragt werden sollen (Jurecka & Hartig, 2007, S. 37). Beim computerbasierten Assessment werden die Aufgaben am Computermonitor präsentiert, die Eingabe der Antworten erfolgt über die Tastatur und Maus bzw. über einen Touch Screen. Die Reaktionen werden elektronisch aufgezeichnet und meist auch elektronisch bzw. automatisiert ausgewertet sowie rückgemeldet.

Bezüglich des Technologieeinsatzes kann ein hierarchischer Bezug hergestellt werden. Die folgende Darstellung veranschaulicht die Beziehung verschiedener einsetzbarer Technologien im Assessment:

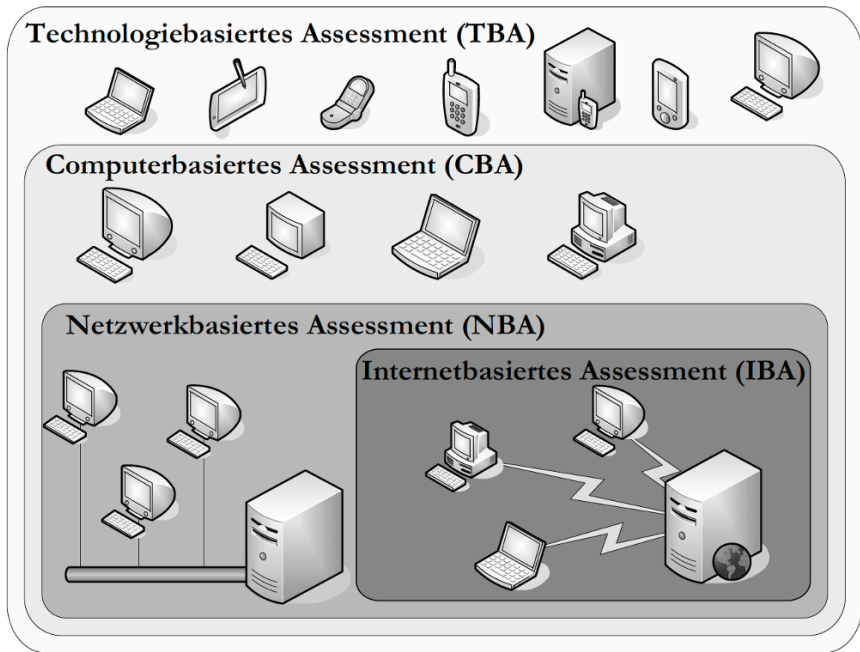


Abbildung 1: Formen des technologiebasierten Assessments (Jurecka & Hartig, 2007, S. 41)

Durchgeführt werden computerbasierte Assessments in einem online oder in einem „blended mode“. Zu computerbasierten Assessments zählen zudem internet- und/oder netzwerkbasierte Assessments. Die Unterscheidung erfolgt anhand der eingesetzten Technologie: Netzwerkbasiertes Assessment bezeichnet die Verwendung von Computernetzwerken (Jurecka & Hartig, 2007, S. 39). Diese können, müssen aber nicht unbedingt über das Internet vernetzt sein. Beim netzwerkbasierten Assessment können Tests an mehreren Computern gleichzeitig mit der Vorgabe von Testbatterien durchgeführt werden. Eine Möglichkeit dieser Form ist die Bearbeitung von Aufgaben in Gruppen. Meistens ist für ein netzwerkbasiertes Testen die Installation von Software notwendig. Im Gegensatz dazu ist die zusätzliche Installation von Software neben der Standardausstattung eines PCs oder Laptops nicht notwendig. Internetbasierte Assessments werden über das Internet und mit einem Browser durchgeführt. Der Vorteil dieser Erhebungsform ist, dass viele Teilnehmer und Teilnehmerinnen mit wenig Mehraufwand (wenn der Test einmal programmiert ist) erreicht werden können. Hier können auch Rollenspiele zur Messung der Teamfähigkeit eingesetzt werden.

Beispielsweise hat IBM Manager in ihrer Teamfähigkeit getestet und geschult, indem diese das Online-Rollenspiel World of Warcraft absolvierten (IBM, 2007). Problematisch ist allerdings die Durchführung von Leistungstests, da internetbasierte Assessments meist in unkontrollierten Umgebungen stattfinden.

Eine Unterform des computerbasierten Assessments ist die computerbasierte Diagnostik. Das Unterscheidungsmerkmal ist hierbei nicht die Technologie, sondern das Ziel der Messung, nämlich die Diagnostik. Unter der computerbasierten Diagnostik werden verschiedene Diagnoseprozesse subsumiert. Computerbasiert diagnostiziert wird beispielsweise im pädagogischen, im pädagogisch-psychologischen und im medizinischen Bereich. Für den pädagogischen Bereich werden demnach Computertechnologien für die Feststellung von Voraussetzungen und Bedingungen planmäßiger Lehr- und Lernprozesse, die Ermittlung von Lernergebnissen sowie die Analyse von Lernprozessen eingesetzt (vgl. zum Begriff Pädagogische Diagnostik Kap. 2.1.1). Ein Beispiel für den pädagogischen Bereich ist der CITO Sprachtest⁷, der in Bremen zur Sprachstandsmessung von Vorschulkindern eingesetzt wird. Dieser erhebt den Anspruch, auf der Grundlage seiner Ergebnisse, Lehrkräfte bei der Einschätzung der Schuleignung zu unterstützen⁸. Auch PISA wird seit der Erhebung 2006 zum Teil computerbasiert durchgeführt. In der Medizin werden 3D-Simulationen und interaktive Itemformate (vgl. Kap. 2.2.5) genutzt, um die diagnostische Kompetenz von angehenden Medizinerinnen und Medizinerinnen zu fördern. Die University of Southern California führt z. B. das Projekt *Virtual Patient* durch, in welchem authentische interaktive Aufgabenformate eingesetzt werden, um klinisches Personal sowohl hinsichtlich medizinischer Diagnostik als auch Kommunikationsfähigkeiten zu trainieren⁹. Wie aus diesen Beispielen hervorgeht, bietet die computerbasierte Diagnostik Möglichkeiten (z. B. 3D-Simulationen), die über den Einsatz papierbasierter Methoden hinausgehen. In den folgenden Kapiteln werden die Anreicherungsmöglichkeiten durch Multimedia und Beispiele für computerbasierte Diagnostik vorgestellt.

⁷ www.de.cito.com (zuletzt geprüft am 03.11. 2014)

⁸ Wie aus einem Bericht der Senatorischen Behörde Bremens hervorgeht, ist es aufgrund mangelnder Computerkompetenzen (insbesondere im Umgang mit der Maus) allerdings nicht allen Kindern möglich den Test zu absolvieren (Die Senatorin für Bildung Wissenschaft und Gesundheit, 2011, S. 5). Auch wenn eine Sprachstandsfeststellung zum Zwecke der Förderung von Sprachberatern/Sprachberaterinnen und Erziehern/Erzieherinnen befürwortet wird, wird der CITO Sprachtest mehrheitlich als ungeeignet bewertet (Die Senatorin für Bildung Wissenschaft und Gesundheit, 2011, S. 8). Somit ist dieser Test und die an ihm geübte Kritik ein Beispiel dafür, dass bei der Entwicklung und dem Einsatz von Erhebungsinstrumenten die Eigenschaften potenzieller Nutzer und Nutzerinnen unbedingt zu berücksichtigen sind.

⁹ <http://ict.usc.edu/prototypes/virtual-patient/>

2.2.2 Rich E-Assessment¹⁰

„E-Assessments¹¹, die interaktive Itemformate einbinden und/oder mit auditiven und visuellen Unterstützungsfunktionen angereichert sind, werden als Rich E-Assessment bezeichnet“ (Wolf u. a., 2011, S. 129).

Der Begriff *Rich E-Assessment* hat seinen Ursprung in dem Begriff *Rich Internet Application*. *Rich Internet Application* bezeichnet Internetanwendungen, in denen Verhaltensmöglichkeiten und Funktionen von Internetanwendungen kombiniert werden und die damit aus vielfältigen Interaktionsmöglichkeiten bestehen (Colombo-Mendoza, Alor-Hernandez & Rodriguez-Gonzalez, 2012, S. 361). Erstmals verwendete Sally Jordan im Jahr 2009 eine Abwandlung des Begriffs Rich E-Assessment: Sie schrieb im Kontext von Potenzialen computerbasierter Assessments von „rich interactive e-assessment“ (Jordan, 2009). Allerdings bezog sie den Begriff ausschließlich auf interaktive Items. Merkmal von einem Rich E-Assessment ist jedoch auch die Anreicherung der Itemformate mit multimedialen Hinweisen, sogenannten *Cues* oder *Prompts* (vgl. Wolf u. a., 2011). Die visuellen oder auditiven Cues dienen der Unterstützung der Orientierung und der Navigation. Zudem kann dadurch die Aufmerksamkeit und Motivation gesteigert werden. Durch diese Unterstützungsfunktionen ist es insbesondere für Personen mit einer niedrigen ICT-Literacy möglich, an Erhebungen in unkontrollierten Umgebungen und ohne personale Unterstützung teilzunehmen (Boyle & Hutchison, 2009, S. 306).

Die Entwicklung und Etablierung von Rich E-Assessment gründete sich u. a. in den Herausforderungen, die durch die Übertragung von einer papierbasierten in eine elektronische Version entstanden. In den Anfängen des E-Assessments wurden meist papierbasierte Tests mit ihren Itemformaten 1:1 in eine elektronische Version übertragen. Dabei zeigte sich, dass die Art des Befragungsmediums einen Einfluss auf die Inhaltsvalidität haben kann: Einerseits ist zu hinterfragen, in wie weit das Itemformat direkt übertragen werden kann und den Messgegenstand evtl. verändert. Andererseits ist zu berücksichtigen, dass besondere Charakteristika – wie z. B. die ICT-Literacy – einen Einfluss auf das Messergebnis haben können (vgl. auch Kap. 2.5 zur Äquivalenzproblematik). Da E-Assessments oft auch als Self-Assessments angeboten werden, sind zudem

¹⁰ Dieses Kapitel wurde in Teilen bereits in Wolf, Koppel & Schwedes (2011) veröffentlicht.

¹¹ Da es sich um einen englischen Begriff handelt, für den es im Deutschen keine äquivalente Übersetzung gibt, wird in diesem Kapitel der englische Begriff Rich E-Assessment verwendet, wobei E-Assessment hier äquivalent mit computerbasierter Diagnostik zu verstehen ist.

motorische und physische Voraussetzungen zu berücksichtigen. Beispielsweise können Seh- und Hörfähigkeit sowie motorische Fähigkeiten einen Einfluss auf das Testergebnis haben. Diese Aspekte werden im Kontext der Äquivalenzproblematik thematisiert (vgl. Kap. 2.5). Allerdings bieten Rich E-Assessments auch gerade dahingehend Vorteile, dass Personen mit eingeschränkten physischen Fähigkeiten durch den Einsatz von Multimedia unterstützt werden können, beispielsweise durch auditive Inhalte sowie die Anpassungsmöglichkeiten der Lautstärke und Darstellungsgröße. Im Folgenden werden die Vorteile von E-Assessments im Allgemeinen und von computerbasierter Diagnostik im Besonderen erläutert.

2.2.3 Vorteile computerbasierter Diagnostik

„Technologies are well suited to supporting many of the data collection, complex analysis, and individualized feedback and scaffolding features needed for the formative use of assessment“ (Quellmalz & Pellegrino, 2009, S. 77).

Durch die technische Entwicklung unterscheiden sich inzwischen die Formate in der computerbasierten Diagnostik stark von den papierbasierten und früheren computerbasierten Formaten. Wie von Bennett erläutert und in einem späteren Artikel mit Csapó erneut erwähnt wurde, verbessert die Technologie die Diagnostik primär auf zwei Weisen: Sie verändert den Prozess an sich und sie erweitert die Messgegenstände selbst (Bennett, 2001; Csapó, Ainley, Bennett, Latour & Law, 2012). Die Verbesserung des Prozesses bezieht sich primär auf ressourcenbezogene Argumente. Die Veränderung der Messgegenstände bezieht sich auf die, mit der Weiterentwicklung der Technik auch ansteigenden, Möglichkeiten zu messender Konstrukte, so z. B. die Messung komplexer Problemlösefähigkeit mit Hilfe von Animationen und Schaubildern oder die Messung von Teamfähigkeit anhand von Rollenspielen.

Daraus ergeben sich folgende Vorteile der computerbasierten Diagnostik (in Anlehnung an Csapó u.a. (2012) sowie Pachler u. a. (2009)):

1. Flexibilität: Die Entwicklung von Tests kann mit wenig Aufwand durch einfache und/oder automatisierende Generierung von Fragen, die einfach modifiziert, verteilt und begutachtet werden können.
2. Einbindung von Multimedia: Es können dynamische Stimuli wie Audio, Video und/oder Animationen eingebunden werden, die den Einsatz von zusätzlichen Medien, wie Audioabspielgeräten, obsolet werden lassen. Zudem können moderne Technologien die Repräsentation von Ideen und die Bearbeitungen der Ideen ermöglichen und Lernende bei dem

Prozess, ihre Ideen zu präsentieren, unterstützen. Die Repräsentationsmöglichkeiten fördern wiederum die Auseinandersetzung mit den Inhalten und öffnen zudem neue Wege der Ideenfindung und –formulierung.

3. Automatisierte Auswertung und Kommunikation: Die Technologie ermöglicht die automatisierte Auswertung von Tests, die Skalierbarkeit sowie Adaptivität. Auch offene Antwortformate können automatisch ausgewertet werden. Die Ergebnisse können unmittelbar rückgemeldet werden, ohne die Notwendigkeit des Ausdrucks oder auch des Verfassens und Verschickens von Emails. Die Auswertung ist individuell oder gruppenbezogen möglich.
4. Datenspeicherung: Es können große Datenmengen gespeichert werden, die Vergleiche zwischen großen Personengruppen und/oder Längsschnittuntersuchungen erleichtern.

Die Vorteile liegen somit zum einen in der Reduzierung der Zeit- und Geldressourcen. Zum anderen in der Automatisierung und in den Gestaltungsmöglichkeiten: Es können authentische Formate erstellt werden, die u. a. eine Identifikation (deren Effekte die Motivationssteigerung und Erhaltung der Aufmerksamkeit sein können) und die Überprüfung komplexerer Fähigkeiten ermöglichen. Dies führt wiederum zur Steigerung der ökologischen bzw. externen Validität.

Trotz aller Vorteile bestehen aber auch Herausforderungen: Diese liegen in der Datensicherheit, dem Programmieraufwand, der Äquivalenz zwischen papier- und computerbasierten Befragungen und dem fraglichen Einfluss der ICT-Literacy (vgl. auch Kap. 2.5 zur Äquivalenzproblematik). Oftmals besteht eine Diskrepanz zwischen der notwendigen Erfassung personenbezogener Daten und dem Anspruch der Datensicherheit.

Speziell für die Umsetzung förderdiagnostischer Prinzipien kann die computerbasierte Diagnostik mit ihren Vorteilen gewinnbringend genutzt werden, worauf im Folgenden eingegangen wird.

2.2.4 Potenziale computerbasierter Diagnostik für die Förderdiagnostik

Unter einer computerbasierten Förderdiagnostik ist die Nutzung von Informations- und Computertechnologie für die Analyse von Kompetenzen unter Einbeziehung des bisherigen Lernverlaufs und der Identifikation von Stärken, Schwächen und der Ableitung von Fördermöglichkeiten zu verstehen (vgl. zum Begriff Förderdiagnostik Kap. 2.1.2). Somit kann die Technologie nicht genuin förderdiagnostischen Prinzipien gerecht werden – aber fast jegliche Technologie kann formativ bzw. förderdiagnostisch genutzt werden (Pachler u. a., 2009, S. 2).

Hinsichtlich der sechs Prinzipien Individualität, Prozessorientierung, Wechselverhältnis von Diagnose und Intervention, Einbettung in das reale Umfeld der Teilnehmenden, Berücksichtigung von Stärken und Schwächen sowie Anwen-

dung von Beobachtungsverfahren und Fehleranalysen (vgl. Kap. 2.1.2) lassen sich die Vorteile einer computerbasierten Förderdiagnostik im Vergleich zu einer papierbasierten Diagnostik folgendermaßen integrieren (in Anlehnung an Pachler u.a. 2009 und Kap. 2.2.3):

Bezüglich der *Individualität* bietet eine computerbasierte Förderdiagnostik die Möglichkeit der Adaptivität und eines unmittelbaren automatisierten Feedbacks. Durch die Nutzung von automatisierten Auswertungsprozessen können die Aufgaben automatisch in Abhängigkeit von bereits gelösten Aufgaben entsprechend des Kompetenzniveaus zugewiesen werden. Auch ist eine individuelle Anpassung von Inhalten (z. B. durch die Auswahl von Aufgaben in bestimmten Themenbereichen) möglich. Ebenfalls durch die automatisierte und unmittelbare Auswertung kann direkt im Anschluss der Bearbeitung einer Lernaufgabe am Computer eine Rückmeldung zur Performanz erfolgen und es können passende Übungsaufgaben zugewiesen werden. Für eine individuelle Förderung kann beispielsweise das automatische Aufrufen eines Regelkatalogs eine Unterstützung darstellen, der bei falscher Setzung von Kommata die entsprechenden Regeln aufzeigt. Auch ist ein sog. *Scaffolding*, bei dem Anleitungen, Denkanstöße und Hilfestellungen gegeben werden, um bei der Lösung von (Teil-) Aufgaben zu unterstützen, möglich (Schnotz, 2006, S. 49). Inzwischen existieren zahlreiche Studien und Einsatzszenarien über das sog. *computer embedded prompting* – das durch einen Reiz ausgelöste Erscheinen von Zusatzinformationen (Cummins, Ardeschiri & Cohen, 2008; Raes, Schellens, De Wever & Vanderhoven, 2012; Sharma & Hannafin, 2007). Die Studien belegen, dass durch den Einsatz von Scaffolding Zusatzinformationen (z. B. nach Nachschlagen von Wörtern) effektiver akquiriert werden können. Aufgrund eines geringeren Aufwands sinken zudem die Hemmschwellen, Zusatzinformationen zu nutzen (Cummins u. a., 2008, S. 17). Ferner unterstützt das computerbasierte Scaffolding nach einer Studie von Reas u.a. die Anreicherung von domänenspezifischem Wissen, das metakognitive Bewusstsein und die metakognitiven Regulationsstrategien. Allerdings ist dennoch zu erwähnen, dass der Lernerfolg mit einem durch Lehrende unterstützten Scaffolding höher ist (Raes u. a., 2012). Dies gilt insbesondere für benachteiligte Lernende, insbesondere jene mit geringem Vorwissen (Kim & Hannafin, 2011). Tendenziell geben Studien zum computerunterstützten Scaffolding Hinweise darauf, dass eine Kombination aus der Unterstützungen durch Lehrende *und* computerbasiertem Scaffolding am effektivsten für das Lernen ist (McNeill & Krajcik, 2009; Puntambekar & Kolodner, 2005; Tabak, 2004). Mit dieser Kombination können auch spezielle individuelle Bedarfe, die ggf. nicht durch ein Scaffolding abgedeckt werden können, adressiert werden.

Die *Prozessorientierung* kann computerbasiert durch automatisierte Auswertungsverfahren und die systematische Speicherung der Performanzwerte

gewährleistet werden. Das Ergebnis der Diagnostik wird gespeichert und mit Ergebnissen weiterer Durchläufe verglichen. So kann der Lernverlauf abgebildet und den Nutzern und Nutzerinnen rückgemeldet werden. Es kann berechnet werden, ob sich eine Person verbessert hat und wenn ja, in welchen Bereichen bzw. in welchem Umfang. Papierbasiert wären die drei Schritte Erhebung, Auswertung und Vergleich manuell vorzunehmen. Das hätte sowohl die Beanspruchung höherer zeitlicher Ressourcen zur Folge als auch ein höheres Fehlerrisiko. Lernstandserhebungen können somit zu mehreren Zeitpunkten durchgeführt, automatisch berechnet bzw. verglichen und den Teilnehmenden rückgemeldet werden, ohne einen hohen personalen Aufwand zu erfordern. Des Weiteren kann die Rückmeldung – insbesondere wenn sie stärkenorientiert erfolgt – zu einer Motivationssteigerung und einem größeren Lerneffekt führen (Jacobs, 2008). Die automatisierte Auswertung ermöglicht zudem eine schnelle Rückmeldung und unterstützt Lernende dabei, ggf. den nächsten Schritt unmittelbar einzuleiten, um im Problemlöseprozess voranzukommen.

Ein *Wechselverhältnis von Diagnose und Intervention* schließt an das Potenzial zur Prozessorientierung an: Computerbasiert kann eine differenzierte Rückmeldung erfolgen, die eine passgenaue Ableitung von Fördermaßnahmen ermöglicht. Eine computerbasierte Diagnostik und Förderung kann zudem adaptiv erfolgen: Die Lern- oder Diagnostikaufgaben werden je nach Performanz der vorherigen Aufgabe an das Niveau angepasst (vgl. z. B. Moosbrugger, 2008, S. 241). Des Weiteren ist eine schnelle Kommunikation mit sowohl Einzelpersonen als auch kleinen oder großen Gruppen sowie die Dokumentation des Kommunikationsprozesses möglich.

Die *Einbettung in das reale Umfeld* kann auf zwei Weisen realisiert werden: Einerseits besteht das Potenzial von neuen Medien insgesamt darin, authentische Aufgabenformate abbilden zu können (Wolf u. a., 2011). Somit kann das Umfeld der potenziellen Zielgruppe in die computerbasierte Förder- und Diagnoseinstrumente durch Simulationen (z. B. Bilder, Videos, Audio) einbezogen werden. Andererseits können mobile Endgeräte genutzt werden, die zeitlich und räumlich flexibel einsetzbar und somit an das individuelle Umfeld sowie den individuellen Tagesablauf angepasst werden können.

Die *Berücksichtigung von Stärken und Schwächen* verlangt durch die automatisierte Auswertung keinen Mehraufwand. Zeitgleich können sowohl die Kompetenzen als auch der Förderbedarf ausgewertet und dargestellt werden. Je nach Zielgruppe oder Bedarf können die Stärken und/oder die Schwächen fokussiert werden: Einerseits, indem die Rückmeldung entweder an den Bedarf angepasst ist oder die Ergebnisse von den Personen selbst ihren individuellen Bedürfnissen entsprechend abgerufen werden. Andererseits können die Diagnose-

und Lernmaterialien je nach gewünschter Fokussierung und Intensität ausgewertet werden.

Für die Anwendung von *Beobachtungsverfahren und Fehleranalysen* bietet die computerbasierte Diagnostik nur bedingt einen Vorteil. In Anlehnung an die Art und die Häufigkeit der Fehler können z. B. automatisch Rückschlüsse über systematische Fehler gezogen werden. Voraussetzung hierfür ist allerdings zum einen, dass die Diagnostik kompetenz- und lerntheoretisch fundiert ist, zum anderen, dass die möglichen Fehler bei der Diagnostik systematisiert und an das dahinter liegende theoretische Modell gekoppelt sind. Um beispielsweise das Nutzer-/Nutzerinnenverhalten hinsichtlich des Gebrauchs von Webseiten zu beobachten, können Beobachtungsverfahren mit Hilfe der Computertechnologie durch das sog. *Screenrecording* unterstützt werden, bei dem der Monitor und somit die Interaktionen zwischen Mensch und Computer aufgezeichnet werden. Die Auswertung kann teilautomatisiert erfolgen. Beim Mouse-Tracking wird beispielsweise die Bewegung der Computermaus aufgezeichnet und hinsichtlich der Zeit (Zeitspanne zwischen Interaktionen), der Verweildauer (wie lange der Mauszeiger auf einer bestimmten Stelle positioniert ist) oder auch der Häufigkeit (wie oft bestimmte Stellen bzw. Schaltflächen aktiviert werden) automatisiert ausgewertet. Ein weiteres Beispiel ist das sog. *Eye-Tracking* (vgl. Kap. 2.7.3). Tiefergehende Analysen können jedoch nur bedingt mit technologischer Unterstützung erfolgen. Beobachtungsverfahren, die komplett automatisiert ausgewertet werden, sind noch recht selten, da die Entwicklung von Algorithmen und deren Zuordnung zu Bedeutungsstrukturen aufwändig sind.

Ausblickend wurde in bisherigen Studien und Überlegungen zum computerbasierten Testen bisher wenig hinterfragt, ob die Art der Gestaltung auch einen Effekt auf das Lernen haben kann. Auch wenn beispielsweise szenariobasierte Testumgebungen eine motivationssteigernde Wirkung nachgesagt wird (s. oben) gibt es bisher wenig Belege dafür, dass computerbasierte Diagnostik nicht nur die Kompetenzen misst, sondern auch das Lernen positiv beeinflusst. Das mag einerseits daran liegen, dass die Feedbackfunktionen teilweise noch nicht ausgereift genug bzw. noch nicht vergleichbar mit menschlichen Rückmeldungen sind. Andererseits realisieren die genannten literalitätsbezogenen computerbasierten Programme zwar eine automatisierte Rückmeldung (ggf. basierend auf einem theoretischen Modell), jedoch ist diese Rückmeldung nicht immer mit einem empirisch überprüften Kompetenzmodell verknüpft. Das bedeutet, ein Nutzer/eine Nutzerin bekommt zwar eine Rückmeldung darüber, ob die Eingabe richtig oder falsch ist, jedoch keine Rückmeldungen hinsichtlich seiner/ihrer (noch zu erweiternden) Kompetenzen.

Das National Research Development Centre for adult literacy and numeracy (NRDC) in England hat einen ausführlichen Report über bestehende literalitäts-

bezogene E-Assessments (und teilweise computerbasierter Diagnoseprogramme) in England vorgelegt. Sie kommen zu dem Ergebnis, dass bestehende E-Assessments hinsichtlich ihrer Gebrauchstauglichkeit (engl. *usability*) unzureichend sind. Insbesondere gilt dies für E-Assessments, die primär von Personen mit einer niedrigen Computerkompetenz (ICT-Literacy) genutzt werden (Brooks u. a., 2005, S. 21). Folgen können mangelnde Motivation, ein dem Lernen hinderlicher *Overload* des Arbeitsgedächtnisses (vgl. Kap. 2.4) oder eine hohe Abbruchquote sein.

Zusammenfassend bietet die Computertechnologie insbesondere durch die Mobilität, die ressourcenschonende Möglichkeit automatisierter Rechenschritte und der Abbildung komplexer Inhalte Vorteile gegenüber papierbasierten Methoden. Wird eine computerbasierte Diagnostik auch als Selbstdiagnostik angeboten, besteht darüber hinaus die Möglichkeit, Hemmschwellen abzubauen: Nutzer und Nutzerinnen können in einem anonymen Umfeld ihre Kompetenzen und Fähigkeiten testen und automatisiert eine Rückmeldung erhalten. Dies kann dazu beitragen, Hemmschwellen zu senken, denn eine Testsituation in einem nicht anonymisierten Umfeld führt wiederum womöglich zu Schamgefühlen hinsichtlich der Offenbarung der eigenen Kompetenzen (vgl. z.B. Döbert u. a., 2000; Egloff, 1997; Füssenich, 2004; Schladebach, 2007). Die genannten Vorteile bezüglich der Ressourcenschonung sollen nicht suggerieren, dass eine betreuende und lehrende Person überflüssig ist. Präsenzsituationen bieten den Vorteil, flexibler auf die Teilnehmenden reagieren zu können. Zudem kann eine Präsenzsituation das Gefühl der sozialen Eingebundenheit stärken und somit zu einer Motivationssteigerung führen (Deci & Ryan, 1993; Friedrich & Mandl, 1997).¹² Zweifellos hängen aber die Vor- und Nachteile von Präsenzsituationen bzw. einem anonymen Umfeld von Testinhalt, Testsituation und der Zielgruppe ab. Es wäre wünschenswert, in Zukunft die Überlegungen über mögliche Lerneffekte durch computerbasierte Diagnostik und über eine zielgruppenspezifische Gestaltung stärker in die Forschung einzubeziehen, um die Potenziale umfassender ausschöpfen zu können.

2.2.5 Beispiele computerbasierter Diagnostik

Insbesondere in den USA ist die Entwicklung von E-Assessment-Techniken weiter vorangeschritten als in Deutschland. Im Folgenden werden Beispiele vorgestellt, in denen die Vorteile computerbasierter Diagnostik – *Flexibilität, Einbindung von Multimedia, automatisierte Auswertung und Datenspeicherung*

¹² Weiterführend sei an dieser Stelle auf Vor- und Nachteile von E-Learning und Blended Learning hingewiesen. Diese geben weitere Hinweise auf lernunterstützende und motivationssteigernde Faktoren von Präsenzsituationen im Vergleich zu mediengestützten Formaten ohne Präsenzformen (vgl. z. B. den Sammelband „Online Lernen“, herausgegeben von Klimsa, 2011).

(vgl. Kap. 2.2.3 zu Vorteile computerbasierter Diagnostik) – zum Ausdruck kommen.

Flexibilität: Ein Beispiel für die Flexibilität ist das e-asTTIE-Tool¹³. Dabei handelt es sich um ein Online Assessment Tool, welches die Messung von Schüler-/Schülerinnenleistungen sowie deren Entwicklung in Lesen, Mathematik, Schreiben und einigen wenigen Fremdsprachen ermöglicht. Lehrende können mit dem Tool die Tests selbst entwickeln, indem auf der Grundlage einer sogenannten linearen heuristischen Programmierung die passenden Itemformate gewählt und an das Curriculum und die Bedürfnisse der Lehrenden angepasst werden können. Die Ergebnisse des Online Assessments ermöglichen die Ableitung individueller Fördermaßnahmen. Die Darstellung des Lernprozesses trägt zum Verständnis der individuellen Lernentwicklung bei. Zudem können Lehrende sowohl eine fachliche Bezugsnorm wählen, welche die individuellen Leistungen mit den Anforderungen im Curriculum, als auch eine soziale Bezugsnorm, welche die Leistungen mit dem nationalen Leistungsdurchschnitt vergleicht, in Beziehung setzt. Alleinstellungsmerkmal ist neben der Möglichkeit des Vergleichs auf nationalem Niveau und der eigenständigen Entwicklung von Items auch die Bilingualität: Neben der englischen Sprache bietet das Tool auch die morische Sprache „te reo“ (polynesische Sprache des indigenen Volkes der Māori in Neuseeland) angeboten.

Einbindung von Multimedia: Durch die Einbindung von Multimedia ist inzwischen auch die computerbasierte Messung „höherer Kompetenzen“ (engl. *higher order skills*) möglich, wie die der komplexen Problemlösefähigkeit oder der Teamfähigkeit. Diese werden beispielsweise mit szenariobasierten Diagnoseprogrammen erhoben. Bei der Software *Primum Computerbased Case Simulations* (National Board of Medical Examiners (NBE), 2009) handelt es sich um ein Instrument, mit dem anhand von Szenarien die Problemlösefähigkeit von (angehenden) Medizinern und Medizinerinnen getestet wird. Jedes Szenario ist eine dynamische und interaktive Simulation eines Patienten-/Patientinnenfalls, um diagnostische Kompetenzen, Behandlungsmethoden und Begleitung zu evaluieren. Dies umfasst die Anamnese, die Diagnose und die daraus folgenden Konsequenzen (wie beispielsweise Überweisung zu anderen Fachexperten/-expertinnen). In Abhängigkeit der Handlungen verändert sich der Zustand des Patienten/der Patientin. Begonnen wird jedes Szenario mit der Darstellung des klinischen Settings, die Angabe der Zeit sowie einführende Informationen zum Patienten/zur Patientin. Die zeitliche Dimension wird ebenfalls berücksichtigt, indem Fälle mit akuten Anliegen sofort behandelt werden müssen und Fälle mit chronischen Leiden auch zeitlich verzögert behandelt können. Verwendet wer-

¹³ <http://e-asttle.tki.org.nz/> (zuletzt geprüft am 03.11. 2014)

den hauptsächlich Multiple Choice- sowie Freitext-Formate. Der Auswertungsalgorithmus wurde auf der Basis von Experten-/Expertinnenurteilen gebildet. Auch die Freitextformate werden automatisch ausgewertet – wobei die Freitextaufgaben nicht zum Schreiben langer Texte, sondern kurzer Berichte auffordern. *Primum* wird eingesetzt für die Prüfungsvorbereitung und auch Teilprüfungen für angehende Mediziner und Medizinerinnen. Für die Nutzung können ausführliche niveaudifferenzierte Manuals herangezogen werden, in denen mögliche Szenarien, Handlungsmöglichkeiten und „Spielregeln“ erläutert werden¹⁴. Bei den Szenarien muss somit nicht nur Wissen angewandt, sondern es müssen komplexe Szenarien analysiert und unterschiedlichste Informationen berücksichtigt werden, die sich in Abhängigkeit der Eingaben der Nutzer und Nutzerinnen verändern.

Im medizinischen Bereich existieren weitere Beispiele, wie die Software *Virtual Patient* der Universität Heidelberg¹⁵. Eines der wenigen Beispiele mit 3D-Simulationen ist das *Virtual Patient Project* von Mitarbeitenden der New York University, bei dem Synergien zwischen Computerspielen und medizinischer Visualisierung genutzt werden¹⁶. Zudem existiert für angehende Neurochirurgen *VCath*, eine *Virtual learning iPad App*, um berufsrelevante Fähigkeiten zu verbessern. Entwickelt wurde sie von Mitarbeitenden der Bangor University¹⁷.

Als weiteres Beispiel computerbasierter Diagnostik der letzten Jahre, in denen der Vorteile der Einbindung von Multimedia deutlich wird, ist das sog. *Recruitment*¹⁸. Dabei handelt es sich um den Einsatz spielerisch-simulativer Elemente im Personalbereich (z. B. bei der Personalauswahl). Der Begriff setzt sich zusammen aus den zwei Begriffen *Entertainment* und *Recruiting* (engl. für Personalbeschaffung). Dabei wird auf die Kombination von spielerischen Ansätzen mit Berufsorientierung, webbasiertem Personalmarketing oder auch mit eignungsdiagnostischer Auswahl via Internet gesetzt, die über die Einbindung von Medien (beispielsweise Videos und Simulationen) realisiert wird. Der Fokus liegt sowohl auf der Auswahl der Kandidaten und Kandidatinnen als auch auf der Informationsvermittlung und Unterhaltung. Für die Auswahl von Bewerbern

¹⁴ Abrufbar sind die Manuals unter <http://www.usmle.org/practice-materials/index.html> (zuletzt geprüft am 03.11. 2014)

¹⁵ <http://www.medizinische-fakultaet-hd.uni-heidelberg.de/> (zuletzt geprüft am 03.11. 2014)

¹⁶ www.tinkering.net/vp (zuletzt geprüft am 03.11. 2014)

¹⁷ <http://www.bangor.ac.uk/cs/full.php.en?nid=15617&tnid=15617> (zuletzt geprüft am 03.11. 2014)

¹⁸ Angemerkt sei, dass Recruitment sowohl als E-Assessment als auch für die computerbasierte Diagnostik eingesetzt wird. Unterscheidungsmerkmal ist dabei der theoretische und empirische Hintergrund: Liegt dem Erhebungs- und Auswertungsprozess ein theoretisch und empirisch fundiertes Modell zugrunde, handelt es sich tendenziell um Diagnostik, wohingegen ein E-Assessment diesen Anspruch nicht verfolgt (vgl. auch Kapitel 2.2 zur Definition).

und Bewerberinnen werden eignungsdiagnostische Inhalte (z. B. kognitive Leistungstests) – unter Berücksichtigung der wissenschaftlichen Gütekriterien Objektivität, Validität und Reliabilität – mit Personalmarketing-Botschaften verknüpft und in einem für die Teilnehmenden möglichst angenehmen Ambiente präsentiert (Kupka, Martens & Diercks, 2011, S. 56).

Die Mitarbeitenden der Firma Cyquest, die sich auf den Bereich Recrutainment spezialisiert hat, erachten die "Selbstausswahl der Kandidaten und Kandidatinnen gegenüber der Fremdauswahl durch die Organisationen, die Verbindung von Serious Games und Social Media als Rekrutierungskanal mithilfe von "Realistic Job Previews" sowie Facebook-Applikationen und die damit verbundene authentische Kommunikation der Arbeitgebermärkte" als Trends des Recruitments (Kupka u. a., 2011, S. 53). Die virtuellen Umgebungen können neben der Unterstützung von Unternehmen zur Auswahl von Bewerbern und Bewerberinnen auch der Selbstdiagnostik dienen. Mit einer Selbstdiagnostik kann im Anschluss an einen Durchlauf dem Teilnehmenden zurückgemeldet werden, wie er/sie im Vergleich zu anderen Teilnehmenden abgeschnitten hat. Über die virtuelle Umgebung und die Fragen wird dem Teilnehmenden zudem ein möglichst authentisches Bild des potenziellen Arbeitsplatzes vermittelt. Die möglichst authentische Darstellung des Unternehmens soll dem Teilnehmenden einen realistischen Eindruck der späteren Anforderungen vermitteln und eine möglichst umfassende Basis für die Entscheidung für oder gegen eine Bewerbung liefern. Wird den potenziellen Bewerbern und Bewerberinnen ein möglichst realistisches Bild vermittelt, können positive Effekte bezüglich der Fluktuation, der Leistungsfähigkeit oder der Abbruchquote erzeugt werden (Kupka u. a., 2011, S. 54). Ein Beispiel für den Einsatz von Recrutainment ist die Kanzlei Houthoff Burama. Die Kanzlei möchte mit dem Recrutainment-Format "Houthoff Burama - The Game" gute Absolventen und Absolventinnen für das Unternehmen interessieren und passende Bewerber und Bewerberinnen identifizieren. In dem Spiel werden potenzielle Bewerber und Bewerberinnen in die Rolle von Anwälten und Anwältinnen bei Houthoff Burama versetzt und dazu aufgefordert, an einem fiktiven aber dennoch realistischen Fall zu arbeiten. Dabei werden allerdings keine juristischen Fachkenntnisse vorausgesetzt, es werden vielmehr Problemlösekompetenzen und soziale Kompetenzen fokussiert.

Voraussetzungen für gelungene Recrutainment-Prozesse sind psychologisch fundierte Anforderungsanalysen sowie deren Operationalisierung und die zielgruppengerechte Gestaltung. Letzteres betrifft sowohl die grafische als auch die technische Umsetzung.

Automatisierte Auswertung: Beispiele für die Entwicklung im Bereich der automatisierten Auswertung sind der sog. *Language Independent Sequence*

Comparison (LISC) und der *Pearson's Versant™ Test*¹⁹. Diese analysieren und bewerten die grammatikalische Struktur von geschriebenen Texten. Die computerbasierte Diagnostik ist weniger zur Literalisierung, als vielmehr für das Lernen einer Zweitsprache gedacht. Der Versant-Test ermöglicht die Messung sowohl gesprochener als auch geschriebener Sprache hinsichtlich linguistischer Einheiten wie Textsegmente, Wörter und Silben. Die Messung der geschriebenen Texte ermöglicht zudem die Auswertung gesamter Sätze. Der LISC ist ein webbasiertes System, mit dem die Übersetzung einzelner Sätze geübt werden kann, indem Fehler erkannt und rückgemeldet werden. Das System vergleicht die eingegebenen Sätze mit möglichen korrekten Antworten hinsichtlich Rechtschreibung und Grammatik (z. B. Verwendung des richtigen Artikels im Deutschen), allerdings wird die Position der Wörter bei der Bewertung nicht berücksichtigt, sondern es wird nur gekennzeichnet, dass sich das entsprechende Wort nicht an einer korrekten Stelle befindet (Fowler, 2008). Weitere Beispiele der automatisierten Auswertung werden in den Kapiteln 2.2.6 und 2.3.6 vorgestellt und näher erläutert.

Datenspeicherung: Der Vorteil der Datenspeicherung kommt durch die prominenten Beispiele der internationalen Vergleichsstudien PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study) und PIAAC (Programme for the International Assessment of Adult Competencies) zum Ausdruck. In PISA wird das sog. TAO-System²⁰ (TAO ist das Akronym für Testing Assisté par Ordinateur, was übersetzt der Ausdruck für computerbasiertes Testen ist) verwendet. Mit Hilfe dieses Systems können große Kohorten parallel getestet und unter bestimmten Kriterien ausgewertet werden. Das System bietet die Umsetzung und das Management des gesamten Prozesses an – von der Itemerstellung bis hin zur Auswertung.

Am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) wird das TAO-System genutzt und weiterentwickelt²¹. TAO unterstützt somit die Möglichkeit der Datenspeicherung von großen Datenmengen, dessen Manage-

¹⁹ www.versanttest.com (zuletzt geprüft am 03.11. 2014)

²⁰ Entwickelt wurde das TAO-System von der Universität Luxembourg sowie dem Centre de Recherche Public Henri Tudor. Weitere Informationen sind abrufbar unter: www.taotesting.com (zuletzt geprüft am 03.11. 2014).

²¹ Am DIPF existiert seit 2007 die Forschungsgruppe „Technology Based Assessment“ (TBA), die grundlagen-, problem- und anwendungsorientierte Forschung zum Thema „Testmethoden für die Zukunft: Innovative Verfahren für die technologiebasierte Kompetenzmessung“ durchführt. Dabei werden die Bereiche computerbasiertes adaptives Testen, computerbezogene Kompetenzen (ICT-Literacy), Effekte technologiebasierter Tests, offene Textformate mit automatischer Antwortcodierung sowie technologiebasierte Diagnostik unter Berücksichtigung von Bearbeitungsschwierigkeiten fokussiert. Eines der Ziele ist, einen internationalen Standard für computerbasiertes Testen zu etablieren.

ment und dessen Auswertung. Es ist zu erwarten, dass durch die Komprimierung von Daten immer weniger Speicherplatz für die Datenspeicherung benötigt wird. Damit steigen auch die Möglichkeiten, Testergebnisse großer Kohorten über einen längeren Zeitraum speichern und auswerten zu können.

Die angeführten Beispiele zeigen lediglich einen kleinen Ausschnitt und geben einen Eindruck davon, welche Möglichkeiten in der computerbasierten Diagnostik (und auch Förderdiagnostik) bestehen. Insbesondere die Optionen zur Simulation, zur Darstellung komplexer Probleme oder dynamischen Wechselwirkungen sowie zur automatisierten Auswertung bringen die Vorteile der computerbasierten Diagnostik zum Ausdruck und stellen entscheidende Unterscheidungsmerkmale und den Mehrwert im Vergleich zur papierbasierten Diagnostik dar. In Bezug auf förderdiagnostische Potenziale dieser Beispiele sei angemerkt (wie bereits eingangs kurz erwähnt), dass es sich bei diesen Beispielen nicht genuin um Förderdiagnoseinstrumente handelt. Weder liefern die Online Tools Ergebnisse zu möglichen Lernstrategien, Lernwiderständen oder systematischen Fehlern (vgl. Kap. 2.1.2), noch werden passende Fördermaßnahmen von dem Systemen vorgeschlagen. Somit können die genannten Beispiele zur Umsetzung von Förderdiagnostik beitragen, nicht jedoch insgesamt als Förderdiagnoseinstrumente bezeichnet werden.

Zweifellos wird der Bedarf an computerbasierter Diagnostik – insbesondere durch groß angelegte internationale Vergleichsstudien wie beispielsweise PISA – und an der Forschung zu diesem Thema bestehen bleiben bzw. weiterhin ansteigen. Mit der (Weiter-) Entwicklung von Technologien und Software steigen die Möglichkeiten zur Entwicklung von Testumgebungen und -szenarien. In den bisherigen Kapiteln nicht explizit betrachtet wurde das Items, ein unabdingbarer Bestandteil eines Tests. Mit der technischen Entwicklung steigen somit auch die Potenziale für die Umsetzung von Itemformaten. Itemformate, in denen technische Neuentwicklungen umgesetzt oder auch auf eine neue Weise miteinander verknüpft werden, werden auch als innovative Itemformate bezeichnet (Parshall, Davey & Pashley, 2000) (vgl. folgendes Kap.). Die angeführten Beispiele haben bereits darauf hingedeutet, dass die Umsetzung computerbasierter Diagnostik je nach Messgegenstand und -absicht variiert. In Zusammenhang mit dem Messgegenstand und der Messabsicht stehen die Auswahl und Gestaltung der Itemformate, mit denen die Daten für die Diagnostik erhoben werden. In den vorherigen Kapiteln wurde auf die computerbasierte Diagnostik im Allgemeinen eingegangen, in den folgenden Kapiteln sollen nun im Besonderen die Charakteristiken innovativer Itemformate (die zum Großteil auch in den eben angeführten Beispielen Anwendung finden) vorgestellt und näher erläutert werden.

2.2.6 Innovative Itemformate

Ein Itemformat ist eine Einheit, welche die Wahl einer oder mehrerer Antwortmöglichkeiten erfordert. Die Antwort wird nach festgelegten Regeln bewertet (Haladyna, 2004, S. 41). Ein Item besteht aus den Komponenten Frage/Aufforderung, Bedingungen für die Angabe der Antwort sowie eine Auswertungsprozedur (Auswertungsalgorithmus). Es gibt verschiedene Formen von Items, wie z. B. offene Itemformate (z. B. Freitexte) und geschlossene Itemformate (z. B. Single Choice).

Mit der Weiterentwicklung von Technologien eröffnen sich neue Räume für die Darstellung und Nutzung von Itemformaten. Im Bereich der computerbasierten Diagnostik werden Items, die technologische Möglichkeiten des Testens (nicht nur bezogen auf Visualisierungsmöglichkeiten, sondern insbesondere auch neue Möglichkeiten der Messung sowie Auswertung) nutzen und die nicht mit konventionellen Testverfahren möglich wären, als innovative Itemformate bezeichnet (Parshall u. a., 2000, S. 129). Dabei geht es nicht nur darum, existierende Verfahren mit neuen Technologien anzureichern, sondern die Technologie dazu zu nutzen, um entweder den *Prozess zu verändern* oder aber auch die möglichen *Messgegenstände zu erweitern* (Bennett, 2001, S. 1). Einerseits steigert die Entwicklung von Technologien die Möglichkeiten einer automatisierten Auswertung. Andererseits ermöglichen neue Technologien auch die Messung von Konstrukten, die vorher nicht oder nur sehr aufwendig gemessen werden konnten (Csapó u. a., 2012, S. 153). Innovative Itemformate können sich daher durch die Erweiterung der *Messgegenstände* und/oder der Automatisierung des *Auswertungsprozesses* auszeichnen.

Hinsichtlich der erstmalig möglichen Messung von Konstrukten nennen Csapó u. a. (Csapó u. a., 2012, S. 153) die Beispiele Problemlösen in technologiereichen Umgebungen (Bennett, Persky, Weiss & Jenkins, 2010), analytische und dynamische Aspekte von Problemlöseprozessen (Wirth & Klieme, 2003), Teamwork mit Hilfe von sog. *Situational Judgement Tests* (SJTs)²² (Kyllonen, 2009) sowie dynamische Entwicklungen anhand der Einbeziehung des Prozesses im Kontext des computerbasierten Assessments (Ainley, 2006; Hadwin, Winne & Nesbit, 2005).

Bezüglich des automatisierten *Auswertungsprozesses* erlaubt die computerbasierte Messung inzwischen auch die Auswertung offener Antwortformate. Auf die Messung von beispielsweise Literalität bezogen, können neben Wörtern oder einzelnen Sätzen inzwischen auch gesamte Aufsätze ausgewertet werden (s. Beispiele computerbasierter Diagnostik in Kap. 2.2.5 und Dimensionen innova-

²² Wobei anzumerken ist, dass auch internetbasierte Computerspiele wie beispielsweise World of Warcraft genutzt werden, um die Teamfähigkeit zu prüfen (IBM, 2007).

tiver Itemformate in Kap. 2.2.6.1). Lückentexte oder offene Formate (wie beispielsweise ein Aufsatz) sind momentan den innovativen Formaten zuzuordnen²³, da die Auswertung solcher Formate komplexe Berechnungen erfordert. Der bereits in Kap. 2.2.5 beschriebene *Pearson's Versant™ Tests*²⁴ misst gesprochene Sprache und das Leseverstehen. Der Auswertung liegt ein linguistisches Modell zugrunde, wobei im Detail die Kompetenzbereiche Aussprache, Struktur, Vokabular, Sprachfluss, Verständnis und Interaktion geprüft werden. Die automatisierte Auswertung der Diagnosesoftware wurde auf der Grundlage vieler Daten in Form von gesprochenen Texten (inklusive Nicht-Muttersprachlern, um eine möglichst große Bandbreite an Akzenten und Dialekten berücksichtigen zu können) und deren Auswertung durch Experten und Expertinnen realisiert (Ripley, Tafler, Ridgway, Harding & Hakan, 2009). Ein ähnliches Vorgehen wurde auch bei dem sogenannten *Pearson Reader*²⁵ gewählt, der die automatisierte Auswertung von Aufsätzen ermöglicht. Im anschließenden Kapitel wird das Beispiel der Literalitätsmessung mit innovativen Itemformaten hinsichtlich des Auswertungsalgorithmus⁶ erneut aufgegriffen.

Bereits diese Erläuterung innovativer Itemformate deutet darauf hin, dass diverse Itemformate mit unterschiedlichen Ausprägungen (z. B. offene und geschlossene Formate) und Gestaltungsmöglichkeiten (z. B. durch die Einbindung von Multimedia) existieren. Entscheidend für die Umsetzung und die Auswahl von Itemformaten ist, was gemessen werden soll: Geht es um fachspezifisches Wissen, Teamfähigkeit oder komplexe Problemlösefähigkeit? Abhängig von dem Messgegenstand sind unterschiedliche Itemformate mehr oder weniger geeignet. Zudem bestimmt das Itemformat die Auswertungsmöglichkeiten. Diese können einfach (z. B. Multiple Choice-Formaten) oder komplex (z. B. offene oder dynamische Antwortformate) sein, wobei letztere nur mit hohem Programmieraufwand automatisiert ausgewertet werden können. Studien belegen, dass Itemformate mit hoher Interaktivität, mit der Möglichkeit Konstruktionsleistungen zu erbringen und den Lernprozess selbst zu kontrollieren sowie zu evaluieren, einen motivationssteigernden Effekt haben (Ally, 2004; S. K. Reed, 2006; Svinicki, 1999).

Die Auswahl passender Itemformate für die Messabsicht kann durch Taxonomien (Klassifikationsschemata) unterstützt werden. Hierfür werden im folgenden Kapitel Dimensionen innovativer Itemformate erläutert und anschließend taxiert.

²³ Die Bewertung „innovativ“ ist natürlich immer im zeitlichen Kontext in Abhängigkeit des aktuellen Entwicklungsstands zu betrachten.

²⁴ www.versanttest.com (zuletzt geprüft am 03.11. 2014)

²⁵ www.pearson.com.au/educator/secondary/digital-learning/pearson-reader-20/ (zuletzt geprüft am 03.11. 2014)

2.2.6.1 Dimensionen innovativer Itemformate

In bestehenden Taxonomien werden Itemformate anhand unterschiedlicher Dimensionen klassifiziert. Die Taxonomien unterstützen damit die dem Messgegenstand angemessene Auswahl passender Itemformate und können vordergründig zwei Organisationsschemata zugeordnet werden: *Messung* und *Computertechnologie* (vgl. Wolf u. a., 2011). Taxonomien, in denen die Möglichkeiten der (diagnostischen) Messung – z. B. die Offenheit des Antwortformates sowie die dem Format entsprechende Auswertungsweise – fokussiert werden, sind dem Organisationsschema *Messung* zuzuordnen (vgl. z. B. Taxonomien von Bennett, 1993; Scalise & Gifford, 2006; Snow, 1993). Taxonomien, die vorwiegend Formate hinsichtlich der eingesetzten Computertechnologie unterscheiden, zählen zum Organisationsschema der *Computertechnologie* (vgl. z. B. Taxonomien von Koch, 1993; Parshall & Harmes, 2007).

In bisher bestehenden Taxonomien wird meist nur eines der beiden Organisationsschemata berücksichtigt (eine Ausnahme stellt die Taxonomie von Parshall, Cavey und Pashley (2000) dar). Darüber hinaus werden die Möglichkeiten zur Einbindung von Medien (zum Beispiel Animationen, Filme, Bilder etc.) nur bedingt systematisch integriert.

In einem Literatur-Review von Wolf u. a. wurden aussagekräftige Taxonomien zu Itemformaten (Boyle & Hutchison, 2009; Parshall u. a., 2000; Scalise & Gifford, 2008) gesichtet (Wolf u. a., 2011)²⁶. Auf der Literaturbasis und unter Berücksichtigung aktueller Entwicklungen in der computerbasierten Diagnostik wurde eine Taxonomie entwickelt, in der sowohl das Organisationsschema Messung als auch Computertechnologie integriert sind. Die Taxonomie ist stark an die von Parshall u. a. (2000) angelehnt, welche aus den Dimensionen Itemformat, Antwortaktivität, Medienanreicherung, Interaktivität und Auswertungsalgorithmus besteht. Die Taxonomie von Wolf u. a. (2011) besteht ebenfalls aus fünf Dimensionen, die allerdings nicht deckungsgleich mit der Taxonomie von Parshall u. a. (2000) sind:

²⁶ Das in diesem Kapitel behandelte Thema wurde bereits in leicht veränderter Form in Wolf, Koppel & Schwedes (2011) veröffentlicht. Die hier vorgestellte Taxonomie weicht in Teilen von den bereits veröffentlichten Inhalten ab. Begründet ist dies darin, dass die Items und die Online-Testumgebung weiterentwickelt und die Bewertung der Itemformate an den Entwicklungsstand angepasst wurden.

1. Offenheit des Antwortformats
2. Auswertungsalgorithmus
3. Authentizität
4. Medienanreicherung
5. Interaktivität

In der Taxonomie von Wolf u. a. (2011) ist im Vergleich zu der von Parshall u. a. (2000) die Dimension *Antwortaktivität* nicht explizit berücksichtigt und sie ist um die Dimension der *Authentizität* erweitert. In den Dimensionen sind beide Organisationsschemata berücksichtigt: *Offenheit des Antwortformats*, *Auswertungsalgorithmus* sowie *Authentizität* sind dem Schema Messung zuzuordnen. *Medienanreicherung* und *Interaktivität* kategorisieren hingegen Gegenstände der Computertechnologie. Die Innovativität stellt eine zusätzliche Dimension dar, die zu den eben genannten Dimensionen „quer“ liegt, d.h. jede der fünf Dimensionen weist neben ihren dimensionsspezifischen Ausprägungen zudem eine Ausprägung in der Dimension *Innovativität* auf. Wie im Folgenden deutlich werden wird, sind die fünf Dimensionen in ihrer Umsetzung nicht gänzlich unabhängig voneinander zu betrachten (so beeinflusst beispielsweise die Offenheit des Antwortformats auch den Auswertungsalgorithmus). Damit hat mitunter ebenfalls die Ausprägung der Innovativität in einer Dimension eines Items einen Einfluss auf den innovativen Charakter in einer weiteren Dimension des Items (z. B. kann eine starke Medienanreicherung auch eine höhere Interaktivität zur Folge haben).

Die Dimensionen werden, unter Zuordnung zu den Organisationsschemata Messung und Computertechnologie, im Hinblick auf ihre Ausprägungsmöglichkeiten und auf den Einsatz in computerbasierter Diagnostik erläutert, um im anschließenden Kapitel die Dimension der Innovativität hinzuzuziehen.

Organisationsschema Messung

1. Offenheit des Antwortformats: Die Dimension Offenheit des Antwortformats beschreibt, wie offen die Handlungsspielräume für die Teilnehmenden bei der Eingabe der Antwort sind. Die Dimension ist als ein Kontinuum zu betrachten, welches im Folgenden nach Scalise und Gifford differenziert wird (Scalise & Gifford, 2006). Als „geschlossenste“ Antwortformate gelten *Multiple Choice*-Formate, bei denen nur eine Antwort (z. B. richtig/falsch oder ja/nein) ausgewählt werden kann (Scalise & Gifford, 2006, S. 11). Allerdings ist anzumerken, dass die Kategorie MC-Formate häufig synonym auch für *Single-Choice*-Formate verwendet wird. Bei Single Choice-Formaten kann lediglich eine Antwort, bei Multiple Choice-Formaten können hingegen auch mehrere Antworten

ausgewählt werden (Haladyna, 2004; Scalise & Gifford, 2006) – worauf auch die Übersetzung von *multiple choice* mit „Mehrfachauswahl“ hinweist.

Die nächste Kategorie auf dem Kontinuum Richtung Offenheit sind halboffene Antwortformate. Zu ihnen zählen *Selection-/Identification*-Formate. Diese Kategorie umfasst *Complex Multiple Choice* bzw. *Multiple Answers*. Hier stehen mehrere Antworten zur Auswahl. Die Distraktoren (falsche Antwortmöglichkeiten) bieten dabei das Potenzial, die Ratewahrscheinlichkeit zu minimieren. Diese Formate, wie sie von Scalise und Gifford definiert werden, werden oftmals als Multiple Choice-Formate beschrieben. Abweichend von der vorgeschlagenen Differenzierung von Scalise und Gifford erscheinen die Kategorien Single Choice (an der Stelle von dem Multiple Choice-Formaten nach dem Verständnis von Scalise und Gifford) und Multiple Choice-Format (an der Stelle von *Selection/Identification*) brauchbarer und verständlicher. Denn der Terminus *Multiple* beim Multiple Choice-Format deutet darauf hin, dass mehrere Möglichkeiten zur Auswahl stehen und sinnvollerweise in Abgrenzung zum Single Choice-Format auch mehrere Antworten gewählt werden können.

Das Format *Reordering/Rearrangement* ist ebenfalls halb offen, es bestehen allerdings mehrere Antwortmöglichkeiten bzw. Kombinationsmöglichkeiten. Dabei werden die Teilnehmer und Teilnehmerinnen aufgefordert, Begriffe zu kategorisieren oder etwas an- bzw. zuzuordnen.

Darauf folgt die Kategorie *Substitution/Correction*, in denen ein Wort korrigiert oder aus bestehenden Elementen eine Figur konstruiert werden soll. Die Korrekturmöglichkeiten sind meist festgelegt, z. B. auf die Buchstaben des Alphabets.

Ein weiteres halboffenes Format ist *Completion*, welches dazu auffordert, kurze Antworten einzutragen, Sätze zu komplettieren, Lückentexte auszufüllen oder Matrizen zu ergänzen. Die Antwortmöglichkeiten (z. B. die Möglichkeiten von Schreibweisen) sind dabei vielzählig (je nach Anzahl der Wörter).

Das Format *Construction* beinhaltet Aufgaben, in denen Lösungen konstruiert werden. Die Konstruktion findet allerdings über die Anwendung einer Methode statt, beispielsweise der Erstellung eines Aufsatzes oder einer Concept Map.

Am anderen Ende des Kontinuums befindet sich das offene Format *Presentation/Portfolio*. Bei diesem Format können Präsentationen, Demonstrationen eines Experiments, Diskussionen oder auch Interviews und die Sammlung/Reflexion von Artefakten für ein (Projekt)Portfolio zur Bewertung herangezogen werden.

Geschlossene Antwortformate sind die am meisten verwendeten Formate in der computerbasierten Diagnostik (Stöckberg, 2012, S. 599). Mit geschlossenen Antwortformaten werden meist Wissen und kognitive Fähigkeiten gemessen; die

Auswertung ist objektiv. Die Schlussfolgerungen beziehen sich meist auf eine Wissensdomäne, Fähigkeiten oder beides (Haladyna, 2004, S. 47). Im Kontext der Förderdiagnostik können mit geschlossenen Formaten einfache Basiskonstrukte getestet werden, um „systematische Misskonzeptionen differenziert auf der Teilkonstruktebene zu diagnostizieren“ (Wolf u. a., 2011, S. 131). Offene Formate können für die Messung komplexer Fähigkeiten oder Kompetenzen – z. B. Schreibkompetenzen – genutzt werden.

2. Auswertungsalgorithmus: Der Auswertungsalgorithmus ist anhand zweier Kriterien zu beschreiben: Automatisierungsgrad und Auswertungsschema. Er beschreibt, in wie weit die Auswertung – bei computerbasierten Tests – automatisiert, teilautomatisiert oder nicht automatisiert erfolgt und welches Auswertungsschema angewendet wird.

Hinsichtlich der Automatisierung ist prinzipiell davon auszugehen, dass eine Antwort umso leichter automatisch ausgewertet werden kann, je geschlossener das Aufgabenformat ist. Bezüglich komplexer und offener Aufgabenformate ist die automatische Auswertung nur bedingt bzw. mit hohem Aufwand möglich. Es gibt aber mittlerweile Programme, die vollautomatisierte Auswertungen sowohl von Kurzantworten, (z. B. dem c-rater (Sukkarieh & Stoyanchev, 2009)) als auch von Aufsätzen ermöglichen (z. B. mit dem e-rater (Burstein u. a., 1998)). Allerdings bedarf eine automatisierte Auswertung von Aufsätzen einer umfangreichen Vorarbeit. Für das Training einer solchen Software und die Entwicklung automatisierter Abläufe zur Auswertung sind 200 bis 250 Aufsätze erforderlich, die von Personen „händisch“ hinsichtlich formaler Anforderungen (Grammatik, Vokabular, Sprachauswahl, Rechtschreibung, Struktur und Kohärenz) ausgewertet und in den Auswertungsalgorithmus einbezogen werden (Streeter, Bernstein, Foltz & Donald, 2011, S. 16). Anhand der festgelegten Auswertungskriterien werden die Aufsätze bewertet. Aufgrund der festgelegten Auswertungskategorien können daher auch nur die bereits verwendete Kriterien Anwendung finden. Die Programme zur automatisierten Auswertung offener Antwortformate werden somit anhand von menschlichen Bewertungen „trainiert“. Eine inhaltliche Analyse können sie nicht leisten.

Vorteile einer vollautomatisierten Auswertung sind die Möglichkeiten, adaptiv zu testen, unmittelbare Rückmeldungen einzubinden (vgl. auch Bemerkung zu schnellere Rückmeldung bezüglich der Dimension Medienanreicherung) und individuelle Fördermaßnahmen abzuleiten. Für das adaptive Testen ist allerdings Voraussetzung, dass die Items in ihrer Schwierigkeit hierarchisch (im Idealfall auf empirischer Basis) eingestuft werden, so dass je nach Kompetenzstand die entsprechenden Items zugewiesen werden können. Die beiden zuletzt genannten Möglichkeiten sind insbesondere für die Erfüllung förderdiagnostischer Prinzipien notwendig (vgl. Kap. 2.1.2).

Neben dem Grad der Automatisierung ist das zweite Kriterium vom Auswertungsalgorithmus das Auswertungsschema, welches dichotom, polytom oder komplex sein kann (Parshall & Harnes, 2007, S. 13). Dichotom bedeutet, es gibt nur richtige oder falsche Antworten. Bei polytomen Items ist die Vergabe von Teilpunkten möglich. Dieses Auswertungsschema wird auch als *partial credit* bezeichnet. Dadurch wird zudem eine unterschiedliche Gewichtung von Messkriterien ermöglicht. Komplexe Auswertungsschemata werden herangezogen, wenn eine große Variation bezüglich der möglichen Antworten besteht. Hierzu zählen beispielsweise auch offene Antwortformate wie Aufsätze. Oftmals sind solche Formate kontextualisiert und mit Unterstützung von Stimuli umgesetzt. Die automatisierte Bewertung solcher Formate ist aufwändig umzusetzen, da durch die große Bandbreite möglicher Antworten zumeist unterschiedliche Auswertungsregeln angewendet werden müssen (beispielsweise schlagen sich bei einer Rechtschreibprüfung die unterschiedlichen Regeln auch in dem Auswertungsalgorithmus nieder). Vorteilhaft ist ein komplexer Auswertungsalgorithmus allerdings dahingehend, dass durch die Berücksichtigung verschiedener Regeln (z. B. nicht nur „richtig“ und „falsch“, sondern auch die Abfolge von Entscheidungen und Gewichtungen dieser Entscheidungen) mehr Informationen über die getestete Person eingeholt werden können.

Hinsichtlich eines förderdiagnostischen Anspruchs sind insbesondere dichotome Auswertungsschemata sinnvoll, da diese am einfachsten eine dimensionsreine Erhebung und Bewertung von Kompetenzen ermöglicht.

3. Authentizität: Eine Ausprägung von Authentizität ist nach Bennett (1993) der Grad der Kontextualisierung (Bennett, 1993, S. 2). Gulikers u.a. differenzieren den Ansatz von Bennett weiter aus und beschreiben Authentizität als ein relatives Konzept:

„[...] authenticity of something can only be defined by its resemblance to something else and it is the specification of this something else that is crucial for further discussion about and examination of the concept of authenticity“ (Gulikers, Bastiaens, Kirschner & Kester, 2008, S. 22).

Ihnen zu Folge ist Authentizität nicht objektiv messbar, sondern wird subjektiv wahrgenommen. Der Grad der Authentizität hängt von den folgenden fünf Charakteristika ab: a) der Art der Aufgabe, b) dem physikalischen Kontext, in dem die Befragung eingebettet ist, c) dem sozialen Kontext des Assessments, d) dem Ergebnis bzw. der Form, in welcher das Ergebnis präsentiert wird sowie e) den Bewertungskriterien. Um eine möglichst hohe ökologische Validität zu gewährleisten, sollten in der Gestaltung authentischer Aufgabenformate das Wissen

sowie die Fähigkeiten realer Aufgaben, welche repräsentativ für das jeweilige „Testfeld“ (im beruflichen Kontext z. B. das Arbeitsfeld) sind und kultur- sowie communityspezifische Praktiken abbilden, berücksichtigt werden (J. S. Brown, Collins & Duguid, 1989). Zudem sollten die Aufgaben die Teilnehmenden dazu auffordern, ihr Wissen, ihre Fähigkeiten und Haltungen so einzubringen, wie sie in einem professionellen Kontext/Arbeitsumfeld von entsprechenden Berufstätigen genutzt und eingesetzt werden würden (J. S. Brown u. a., 1989; Darling-Hammond & Snyder, 2000; Gielen, Dochy & Dierick, 2003; J. Herrington & Oliver, 2000). So sollten auch im Sinne einer möglichst hohen prognostischen Validität (verstanden als Voraussage für die zukünftige erfolgreiche, z. B. berufliche Tätigkeit) die Aufgaben weitestgehend der künftigen Tätigkeit entsprechen. Die Bewertungskriterien sind entsprechend der geforderten Ergebnisse in der Praxis zu gestalten (Darling-Hammond & Snyder, 2000). Begründet ist diese Ausdifferenzierung in fünf Charakteristika darin, dass der Grad der Authentizität durch Beeinflussung der einzelnen Charakteristika und den Grad der Kontextualisierung verändert werden kann. Die durch die Authentizität entstandene Komplexität hat wiederum einen Einfluss auf das Frageformat und begründet somit die Zuordnung der Authentizität zum Organisationsschema der Messung. Diese Erläuterung macht die Grenze zur Interaktivität (siehe weiter unten) deutlich: Z. B. kann ein Formular hochgradig kontextualisiert und somit authentisch sein; es fordert aber keine Interaktivität über das Ausfüllen der Formularelemente hinaus. Hingegen kann eine hoch interaktive Simulation vollkommen abstrakt und somit nicht authentisch sein.

Allerdings ist gerade in der erwachsenengerechten Diagnostik aus Motivationsgründen auf die Verwendung eines arbeits- und lebensweltbezogenen Kontext zu achten. Rich E-Assessments (vgl. Kap. 2.2.2) bieten aufgrund der Möglichkeit der Einbindung verschiedener Medien ein großes Potenzial.

Organisationsschema Computertechnologie

4. Medienanreicherung: Die Medienanreicherung beschreibt die Anzahl sowie die Daten- und Informationsdichte der Medienformate, die bei der Gestaltung einer Aufgabe genutzt werden. Sie ist dem Organisationsschema Computertechnologie zuzuordnen. Im Vergleich zur Media-Richness-Theorie²⁷ von Daft und Lengel (Daft & Lengel, 1986) geht es hierbei allerdings nicht um eine Passung zwischen Kommunikationsanlass und Medienwahl sondern vorwiegend um deren Varianz. Nach Dennis und Valacich wird der Reichtum eines Mediums wie folgt beschrieben:

²⁷ In der Media-Richness-Theorie wird der Frage nachgegangen, wie eine effektive Kommunikation und Kooperation durch eine angemessene Medienwahl unterstützt werden kann.

„Richer media were those with a greater language variety (the ability to convey natural language rather than just numeric information), a greater multiplicity of cues (the number of ways in which information could be communicated such as the tone of voice), a greater personalization (ability to personalize the message), and more rapid feedback” (Dennis & Valacich, 1999).

Reiche Medien bieten demnach die Möglichkeiten einer Varianz der *Inhaltsdarstellung* (z. B. Bild- und Textsprache), der *Medieneinbindung* (Bild, Ton etc.), der stärkeren *Personalisierung* sowie der *schnelleren Rückmeldungen* (vgl. hierzu Dimension Auswertungsalgorithmus²⁸).

Inhaltsdarstellung: Inhalte können mit Hilfe der Computertechnologie auf verschiedene Weisen dargestellt bzw. codiert (Weidemann, 2002, S. 47) werden. Unterschieden wird dabei zwischen Text, Bild und Zahlen. Wird der Inhalt nur auf eine Weise dargestellt, handelt es sich um eine monocodale Darstellung – z. B. wenn der Inhalt nur über die Textebene präsentiert wird. Werden hingegen mehrere Darstellungsoptionen genutzt, liegt eine multicodale Inhaltsdarstellung vor. Über die Computertechnologie können Inhalte unterschiedlich und/oder parallel codiert werden. Studien zeigen, dass die Verwendung unterschiedlicher Codierungen je nach Kombinationen mit der Darstellungsmodalität einen positiven oder aber auch negativen Effekt auf den Lernprozess haben können. So können Inhalte beispielsweise besser aufgenommen werden, wenn eine Kombination zwischen Grafik und Audio vorliegt, im Vergleich zu einer gleichzeitigen Präsentation von Grafik und Text (vgl. Low & Sweller, 2005; Weidemann, 2002).

Medieneinbindung: Computer ermöglichen die Einbindung und Kombination unterschiedlicher Medien, wie Grafiken, Audio, Video und Animationen (Parshall u. a., 2000). Grafiken sind dabei die gebräuchlichste Form der nicht textbasierten Darstellung. Auch in papierbasierten Tests könne Grafiken eingebunden werden, doch ist es mit Hilfe von Computern und der Verbindung mit Interaktivität möglich, Grafiken zu bearbeiten, sie rotieren zu lassen und sie zu vergrößern oder zu verkleinern.

Audio wird insbesondere für die Messung musikalischer und sprachlicher Fähigkeiten (Hörverständnis) verwendet. Es können zwar auch Audioabspielgeräte in Verbindung mit einer papierbasierten Diagnostik verwendet werden, doch erfordert dies die zusätzliche separate Nutzung von Abspielgeräten, die aber nicht direkt mit dem Testinhalt verknüpft sind. Auf computerbasierter Basis

²⁸ Hinsichtlich der Medienanreicherung wird die „schnellere Rückmeldung“ nicht weiter berücksichtigt, das sich die unmittelbare Rückmeldung aus dem Grad der Automatisierung ergibt.

können Darstellungen direkt mit Audio verknüpft werden, indem z. B. auf dem Bildschirm ein Symbol oder Bereich mit der Computermaus angeklickt und die dazugehörige Audiodatei abgespielt wird. Auch ermöglicht die Einbindung von Audio die Nutzung von Computertechnologie für Personen mit geringen Lese- oder Sehfähigkeiten. Zudem kann die Einbindung von Audio das Arbeitsgedächtnis entlasten: Das Modalitätsprinzip besagt, dass – im Vergleich zu einer Inhaltsdarstellung in Textform – die auditive Darstellung in Kombination mit einer graphischen Darstellung das Arbeitsgedächtnis entlastet und somit Lernen unterstützt (Low & Sweller, 2005).

Videos ermöglichen die Darstellung von Inhalten, die papierbasiert nur schwer oder gar nicht abbildbar sind, beispielsweise komplexe Sachverhalte, die authentische Darstellung von Prozessen (z. B. Produktionsprozessen) oder von Personen (z. B. im medizinischen Bereich die Darstellung eines Patienten-/Patientinnenfalls). Wie auch die Nutzung von Audio kann Video ebenfalls im Kontext papierbasierter Testung genutzt werden. Doch liegt der Mehrwert der integrierten Nutzung von Computer und Video insbesondere darin, das Video individuell abspielen, anhalten und/oder wiederholen zu können. Gleichzeitig ist aber zu beachten, dass die Decodierung eines Videos, in dem wenig komplexe Inhalte dargestellt werden, eine erhöhte Verarbeitungsleistung im Vergleich zu einer weniger medienreichen Darstellung (z. B. einem Bild) erfordert.

Animationen haben im Vergleich zum Papiermedium den Vorteil, dass sie nicht statisch sind. Sie stellen insbesondere einen Mehrwert durch die Möglichkeit dar, komplexe dynamische Prozesse darzustellen. Animationen können gegenüber Videos vorteilhaft sein, da sie die schematische Abbildung von Prozessen und/oder auch nur einen Teilausschnitt dieser ermöglichen. So ist auch die Fokussierung von bestimmten Aspekten möglich. Des Weiteren ist für die Darstellung von Animationen weniger Rechenkapazität notwendig als für die Darstellung von Videos (Parshall u. a., 2000, S. 139).

Der Stimulus eines Mediums kann sich zudem auf weitere menschliche Sinne beziehen. Mit *tangible Interfaces* ist es möglich, einen haptischen Reiz auszulösen – so werden z. B. in der Medizin Puppen verwendet, in denen der Puls simuliert wird. Auch der Geruchssinn kann mit Hilfe von Geruchssynthesizern stimuliert werden.

Personalisierung: Inhalte, Medien und Rückmeldungen können an die jeweiligen Bedürfnisse angepasst werden. Die inhaltliche Anpassung erfolgt z. B. in adaptiven Tests, indem in Abhängigkeit der Beantwortung eines Items das darauf folgende Item präsentiert wird (vgl. auch Kap. 2.2.3 zu Vorteilen computerbasierter Diagnostik). Die eingesetzten Medien können an die individuellen Bedürfnisse angepasst werden: Bei Audios und Videos durch die Einstellung der Lautstärke, der Möglichkeit der Unterbrechung und/oder der Wiederholung so-

wie bei Graphiken durch die Anpassung der Größe. Die Rückmeldungen können je nach Bedarf beispielsweise hinsichtlich einer Stärken-/Schwächenorientierung, des fokussierten Themengebiets, des Detaillierungsgrades und/oder der gewünschten Bezugsnorm ausgegeben werden. Darüber hinaus können Nutzer-/Nutzerinnenprofile mit persönlichen Angaben erstellt werden, in denen der Lernverlauf abgebildet ist.

Aus der Darstellung zur Mediananreicherung ist weder zu schlussfolgern, dass eine möglichst große Varianz immer sinnvoll ist; noch ist daraus abzuleiten, dass die Ausprägung der Innovation umso stärker ist, je größer die Varianz ist. Die Codierung des Inhalts und die Medieneinbindung sind immer unter Berücksichtigung der Aufgabe und der Zielgruppe zu wählen. Medienreichtum darf nicht zu einer Überforderung der Rezipienten führen. Wie auch in der CLT und der CTML (vgl. Kap. 2.4) beschrieben ist zu vermuten, dass eine zu reichhaltige Darstellung, die mehrere Informationskanäle (z. B. Text, Bild und Ton) beansprucht, so viele kognitive Ressourcen zur Decodierung bindet, dass die eigentliche Lösung der Aufgabe erschwert wird und somit die Konstruktvalidität verschlechtert werden würde. Besonders bezüglich der Zielgruppe der funktionalen Analphabeten und Analphabetinnen ist ein geringer Toleranzbereich zu vermuten (vgl. Kap. 2.3). Die Medienwahl ist an deren Lese- und Schreibkompetenz sowie deren ICT-Literacy anzupassen. So sind die auditive Unterstützung von Instruktionen und die Veranschaulichung von Bildern sinnvoll. Voraussichtlich würden aber zu viele Variationsmöglichkeiten und eine dichte Darstellung von unterschiedlichen Inhalten zu einer *Overcomplication* (Schwabe, 2001, S. 55) führen. Im Idealfall führt der Einsatz von Medien allerdings zu einer Reduzierung von kognitiven Ressourcen, da mehrere Sinneskanäle angesprochen werden und/oder literale Defizite durch auditive Unterstützung kompensiert werden können.

Zusammenfassend liegen die Vorteile einer reichen Medieneinbindung zum einen darin, dass die Herstellung einer oft gewünschten Authentizität und Komplexität nur durch die Mediananreicherung möglich ist. Zum anderen beinhaltet ein hoher Medienreichtum das Potenzial, die Testmotivation aufrecht zu erhalten. Handelt es sich allerdings um rein dekorative Medien, die womöglich nur einen geringen Bezug zum Aufgabeninhalt herstellen, ist mit einer Verschlechterung der Inhaltsvalidität zu rechnen (Laitusis, 2010; Mayer, 2005). Schlussfolgernd ist ein Itemformat bezüglich der Mediananreicherung dahingehend innovativ, wenn das Potenzial neuer Technologien genutzt wird, indem die Medien und deren Kombination im Sinne des Messgegenstands und der Zielgruppe sinnvoll gewählt werden.

Aufgrund dessen, dass sich die Medienanreicherung auf die technische Integration verschiedener Medienarten bezieht, ist diese Dimension dem Organisationsschema Computertechnologie zuzuordnen.

5. Interaktivität: Interaktivität bezeichnet das Aufeinanderbeziehen von wechselseitigen Kommunikations- und Handlungsprozessen von zwei oder mehr aktiven Agenten, die alternierend zuhören, denken und/oder sprechen (Crawford, 2004, S. 29). Dabei werden folgende Aspekte unterschieden: (a) die Schnelligkeit der Reaktion auf die Aktion des Interaktionspartners (z. B. der Klick auf eine Schaltfläche in einem Programm); (b) die Tiefe (beispielsweise der Rechenaufwand eines Schachprogrammes zur Berechnung des nächsten Zuges oder das Handlungsmuster eines computergesteuerten Spielpartners in einem Computerspiel) und (c) die Wahlmöglichkeiten, insbesondere die funktionale Signifikanz einer Entscheidung auf den weiteren Verlauf der Interaktion und die wahrgenommene Vollständigkeit der Wahlmöglichkeiten.

Voraussetzung für die Interaktivität ist, dass die computerbasierte Diagnostik die Handlungen des „Interaktionspartners“ – in diesem Fall der Mensch – eindeutig interpretiert werden, um den Verlauf einer sinnvollen Interaktionskette zu ermöglichen. Demnach sind alle Computerprogramme mit implementierten Feedbackprozessen als interaktiv zu bezeichnen: In Abhängigkeit der Aktionen – der Eingaben – erfolgt eine Rückmeldung des Computers bzw. des Programms. Die Dimension Interaktivität bezieht sich insbesondere auf die programmier-technische Implementation und ist daher ebenfalls dem Organisationsschema Computertechnologie zuzuordnen.

2.2.6.2 Taxonomie innovativer Itemformate

Aus dem vorherigen Kap. geht hervor, hinsichtlich welcher Dimensionen die Itemformate charakterisierbar sind. Eine zu den fünf Dimensionen *Offenheit des Antwortformats*, *Auswertungsalgorithmus*, *Authentizität*, *Medienanreicherung* und *Interaktivität* quer liegende Dimension ist die *Innovativität*. In Abhängigkeit der Dimensionsausprägung verändert sich auch das *innovationspotenzial*. Die folgende Tabelle zeigt eine Taxonomie von innovativen Itemformaten, in der neben den fünf Dimensionen sowohl beide Organisationsschema als auch die Dimension der Innovativität berücksichtigt sind.

Organisationschema	Dimension	Innovationspotenzial		
		niedrig		hoch
Messung	1. Offenheit des Antwortformats	geschlossen	halboffen	offen
	2. Auswertungsalgorithmus Automatisierung Komplexität	nicht automatisiert dichotom	teilautomatisiert polytom	automatisiert komplex
	3. Authentizität	abstrakt	teilkontextualisiert	kontextualisiert
Computertechnologie	4. Medienanreicherung Inhaltsdarstellung Medieneinbindung Personalisierung	monocodal keine Varianz nicht möglich	duocodal wenig Varianz teilweise möglich	multicodal viel Varianz möglich
	5. Interaktivität	niedrig	mittel	hoch

Tabelle 2: Taxonomie innovativer Itemformate (eigene Darstellung in Anlehnung an Wolf, Koppel & Schwedes 2011)

In der linken Spalte sind die Organisationsschemata aufgeführt, denen in der zweiten Spalte die fünf Dimensionen mit ihren Subkategorien zugeordnet sind. Die Dimensionen werden jeweils in ihren extremen und mittleren Ausprägungen qualifiziert. Je stärker ein Item in seiner Ausprägungstendenz zu den in der rechten Spalte aufgeführten Attributen tendiert, als desto größer ist auch das Innovationspotenzial einzustufen.

Der oben aufgeführten Definition von innovativen Itemformaten folgend, bieten offene Antwortformate Möglichkeiten der Messung, die mit geschlossenen Antwortformaten nicht oder nur bedingt möglich sind (z. B. komplexe Problemlösefähigkeit). Die Möglichkeiten zur Abbildung von Komplexität und automatisierter Auswertung ist in konventionellen papierbasierten Tests ebenfalls weniger möglich als in technologiebasierten Umgebungen. Damit einher geht auch die Medienanreicherung und die Interaktivität. Eine multicodale Darstellung von Inhalten mit einer hohen Medienvarianz und Option zur Personalisierung kann die Kriterien eines innovativen Items (Nutzung der Technik, Visualisierungsmöglichkeiten und Erweiterung des Messgegenstands) in einem weitaus höheren Maße erfüllen als es papierbasierte Tests oder einfache Single Choice-Aufgaben am Computer können. Der in der Taxonomie verwendete Begriff *Innovationspotenzial* vermittelt, dass kein unabdingbarer kausaler Zusammenhang zwischen dem Ausprägungsgrad der fünf Dimensionen und der Innovativität

besteht: Eine hohe Medieneinbindung und Authentizität bedingt nicht eine gleichzeitig hohe Innovativität. Im Vordergrund steht immer, die Dimensionen und deren Ausprägungsgrad im Sinne des Messgegenstands und der Zielgruppe zu wählen und miteinander zu kombinieren.

Anhand des folgenden Beispiels wird die Einsatzmöglichkeit der Taxonomie verdeutlicht: Für die Entwicklung einer computerbasierten Diagnostik zur Messung der Schreibfähigkeit bei wenig literalisierten Personen kann die Taxonomie herangezogen werden, um sowohl den Messgegenstand und die -absicht als auch zielgruppenspezifische Charakteristika zu berücksichtigen. Die Schreibfähigkeit kann z. B. über die Rechtschreibfähigkeit von einzelnen Wörtern geprüft werden. Dafür wäre ein offenes Antwortformat (Textfeld oder Freitexteingabe) notwendig. Die Auswertung würde dichotom (richtig/falsch) erfolgen. Eine starke Ausprägung der Authentizität wäre nicht notwendig, da die Rechtschreibung von Wörtern kontextunabhängig messbar ist. Die Misserfolgsbefürchtungen können bei funktionalen Analphabeten und Analphabetinnen hoch sein (vgl. Kap. 2.3), daher ist es unter Umständen hilfreich, authentische Formate einzubeziehen, um die Motivation aufrecht zu erhalten bzw. zu steigern. Für eine Inhaltsdarstellung kann eine duocodale Ausprägung sinnvoll sein, um das Arbeitsgedächtnis zu entlasten und nicht nur Text zu präsentieren, dessen Decodierung, je nach Niveau, bei wenig literalisierten Personen einen enormen Arbeitsaufwand bedeuten kann (Ermüdung und Frustration können die Folge sein). Hinsichtlich der Medienanreicherung ist mindestens eine geringe Varianz sinnvoll, um die Nutzer und Nutzerinnen mit Audiodateien zu unterstützen und die Instruktionen vorlesen lassen zu können. Eine ausgeprägte Personalisierung wäre nicht notwendig; zudem könnte eine zu starke Personalisierung zu einer Überforderung aufgrund der tendenziell niedrigen ICT-Literacy führen. Eine Personalisierung wäre diesbezüglich allerdings gewinnbringend realisierbar hinsichtlich der Abbildung des individuellen Lernverlaufs, indem Lernende einen Fokus setzen und z. B. auswählen können, welchen Kompetenzbereich sie sich anzeigen lassen möchten. Auch die Interaktivität sollte aufgrund der tendenziell niedrigen ICT-Literacy gering gehalten werden. Personen mit einer sehr niedrigen ICT-Literacy können durch hoch interaktive Itemformate leicht überfordert werden.

Zusammenfassend kann diese Taxonomie sowohl für die Entwicklung von Itemformaten und die sinnvolle Abstimmung des Ausprägungsgrades als auch für die Kategorisierung und Bewertung von Itemformaten herangezogen werden. Bei der Entwicklung von Items wird meist von dem zu messenden Konstrukt ausgegangen, wodurch bereits in vielen Fällen eine Präferenz für die Formatauswahl besteht. Die Wahl für die Offenheit des Antwortformats beeinflusst u. U. den Ausprägungsgrad weiterer Dimensionen. Hinsichtlich der personalen

Ressourcen ist die Frage zu stellen, ob die Auswertung automatisiert stattfinden soll oder ob die Auswertung der Aufgaben durch Personen durchgeführt werden kann. Wie erwähnt, können geschlossene Formate im Vergleich zu offenen Formaten leicht automatisiert ausgewertet werden; ein Aufsatz kann nur bedingt automatisiert ausgewertet werden und es bedarf eines hohen Entwicklungs- und Programmieraufwandes (vgl. Kap. 2.2.5 und z.B. Sukkarieh & Stoyanchev, 2009). Vor dem Hintergrund dieser Entscheidungen bzw. Rahmenbedingungen ist bzgl. der prognostischen Validität und der Testmotivation weiterhin zu klären, inwieweit die zu entwickelnden Items authentisch, interaktiv und medienreich sein sollten (Wolf u. a., 2011, S. 139). Dafür ist insbesondere die Zielgruppe zu berücksichtigen. Mit der Taxonomie können die Ausprägungen der Items unter Berücksichtigung der Messabsicht, des Messgegenstands und der Zielgruppe gewählt werden, um eine sinnvolle Umsetzung zu ermöglichen.

2.3 Zielgruppenspezifische Voraussetzungen

Wie bereits in den vorherigen Kapiteln angedeutet, sind Diagnoseverfahren immer kontextgebunden. Das bedeutet, die Verfahren sind an die Zielgruppe und die Rahmenbedingungen anzupassen, indem zielgruppenspezifische Charakteristika und der potenzielle Einsatzbereich berücksichtigt werden. Um ein computerbasiertes Diagnoseinstrument für funktionale Analphabeten und Analphabetinnen entsprechend dieser Zielgruppe entwickeln und einsetzen zu können, sind folgende Kontextvariablen zu berücksichtigen: Literalität, Struktur/Verteilung des Ausmaßes funktionalen Analphabetismus⁷, Herausforderungen in der Alphabetisierungspraxis sowie die Computerkompetenzen von funktionalen Analphabeten und Analphabetinnen. Diese werden im Folgenden dargestellt.

2.3.1 Literalität

Der Begriff Literalität ist auf den lateinischen Begriff *littera* für *Buchstabe* zurückzuführen, was vorerst nur auf die Fähigkeit, mit Buchstaben umgehen zu können, deutet aber keineswegs eine kompetenzbezogene Definition liefert. Eine Definition von Literalität lieferte die UNESCO 1978:

"A person is literate who can with understanding both read and write a short simple statement on his everyday life" (UNESCO, 1978).

Eine weitere Definition wurde von der OECD formuliert: Literalität ist "die Verwendung von gedruckten und geschriebenen Informationen, um in der Gesellschaft zurechtzukommen, eigene Ziele zu erreichen und eigenes Wissen sowie die individuellen Möglichkeiten zu entwickeln" (OECD, 1995, S. 16). Diese

Definition wurde zudem in der IALS-Studie verwendet. In vorangegangenen Diskussionen wurde der Begriff Literalität zunehmend mit Grundbildung synonym verwendet (Grotlüschen, 2011, S. 14). Begründet ist dies womöglich in der Übersetzung von *literacy* im Englischen in *Literalität* ODER *Grundbildung* im Deutschen. So wird auch im deutschen Sprachraum beispielsweise von *computer-* bzw. *ICT-Literacy* gesprochen. Im englischsprachigen Raum finden vor allem auch Begriffe wie *political literacy* Verwendung (Linde, 2007, S. 94).

Deutlich wird, dass in beiden Definitionen (der UNESCO und der OECD) eine Kompetenz- bzw. Fähigkeitsdimension verwendet wird, nämlich "verstehen", "zurechtkommen" und „Entwicklung“. Eine weitere Dimension ist die Gesellschaftliche mit den Beschreibungen "everyday life" und "in der Gesellschaft zurechtkommen". Die gesellschaftliche Dimension impliziert, dass es sich bei diesem Begriff (wie auch beim Begriff "Funktionaler Analphabetismus" - vgl. folgendes Kap.) um ein relationales Konzept handelt. Was Nickel (2011) in Bezug auf Funktionalen Analphabetismus schreibt, kann hier auch auf den Begriff Literalität bezogen werden, "da die Anforderungen an Schriftsprachlichkeit historisch und kulturell wandelbare Größen sind und daher variieren" (Nickel, 2011, S. 54). Internationale Literalitätsstandards (hier verstanden als Lese- und Schreibfähigkeiten) können somit nur anhand "kulturell-funktionaler" Dimensionen definiert werden, die wiederum in Abhängigkeit des gesellschaftlichen Kontextes entsprechend betrachtet werden müssen.

Abhängig vom gesellschaftlichen Kontext wird dementsprechend zwischen literalisierten und nicht bzw. wenig literalisierten Personen unterschieden. Der zuletzt genannten Gruppe gehören die Analphabeten/Analphabetinnen sowie die funktionalen Analphabeten und Analphabetinnen an. Da die funktionalen Analphabeten und Analphabetinnen die Zielgruppe dieser Arbeit darstellen, wird der funktionale Analphabetismus im Folgenden näher erläutert.

2.3.2 Funktionaler Analphabetismus

Funktionaler Analphabetismus ist ein weltweit existierendes Phänomen. Auch in hoch industrialisierten Ländern und Ländern mit einem hohen Bildungsstand tritt das Phänomen auf. Funktionaler Analphabetismus ist ein Begriff, der sich auf die Lese- und Schreibkompetenzen von Erwachsenen bezieht. Eine Definition, die nicht an operationalisierten Sprachstandsbeschreibungen festhält kann somit nur auf einer höheren Ebene erfolgen.

In einem Report zur Grundbildung in Europa wird von der *French National Agency for the Fight against Illiteracy* festgehalten:

"Literacy and numeracy definitions need to be operational rather than global or static and are highly dependent on the actual assessment tool. The com-

plexity of linking definition to literacy level need always be borne in mind and the purposes of those used in testing be quite clear from the earliest stages through to actual use of results" (ANLCI - French National Agency for the Fight against Illiteracy, 2009, S. 7).

Sprachstandserhebungen geben somit immer Werte an, die auf das jeweilige Erhebungsinstrument und die erhobene Sprache referieren und mit dem dahinterliegenden Verständnis von Literalität verknüpft sind. Dennoch bestehen Bemühungen, eine möglichst weitreichende Definition zu etablieren. So hat eine Gruppe von Wissenschaftlern und Wissenschaftlerinnen aus dem Förderschwerpunkt „Forschung und Entwicklung zur Alphabetisierung und Grundbildung Erwachsener“ (Laufzeit 2007-2011) im Jahr 2010 eine Definitionsgrundlage zum Funktionalen Analphabetismus geliefert²⁹:

„ „Funktionaler Analphabetismus“ ist gegeben, wenn die schriftsprachlichen Kompetenzen von Erwachsenen niedriger sind als diejenigen, die minimal erforderlich sind und als selbstverständlich vorausgesetzt werden, um den jeweiligen gesellschaftlichen Anforderungen gerecht zu werden. Diese schriftsprachlichen Kompetenzen werden als notwendig erachtet, um gesellschaftliche Teilhabe und die Realisierung individueller Verwirklichungschancen zu eröffnen.

Unter schriftsprachlicher (literal)er Kompetenz ist die Fähigkeit zu verstehen, sich der Schrift als Kommunikationsmittel zu bedienen. [...] Dies ist gegenwärtig zu erwarten, wenn eine Person nicht in der Lage ist, aus einem einfachen Text eine oder mehrere direkt enthaltene Informationen sinnerfassend zu lesen³⁰ und/oder sich beim Schreiben auf einem vergleichbaren Kompetenzniveau befindet.“

²⁹ Die Definition ist ohne Angabe von Autoren oder Autorinnen auf www.grundbildung.de/daten/Grundlagen/definition (zuletzt geprüft am 03.11. 2014) erschienen.

³⁰ Dies entspricht in der PISA-Studie der Kompetenzstufe 1 („Eine oder mehrere unabhängige, leicht auffindbare Information(en) lokalisieren; Voraussetzung für das Auffinden der Information: wenig konkurrierende Informationen im Text“; vgl. PISA 2000, Opladen 2001, S. 89). In der IALS-Studie

Ihren Ausgangspunkt nimmt diese Definition in der Definition der UNESCO von 1978 und folgt einer Art Baukastenprinzip, indem anfangs das Problem benannt und im weiteren Verlauf näher bestimmt wird. Nach Angabe der Verfassenden können je nach Fragestellung oder Projektkontext die angegebenen Erläuterungen variiert oder ergänzt werden. Auch bei dieser Definition wird sich wie bei der Definition von Literalität einer kulturell-funktionalen Dimension bedient. Die leo.- Level-One-Studie liefert einen Beitrag hin zu einer stärkeren Operationalisierung, indem ein Bezug zu den Kompetenz-Level-beschreibungen (vgl. Kap. 4.1 zum lea.-Projekt) hergestellt wird:

„Die UNESCO spricht von funktionalem Analphabetismus bei Unterschreiten der vollen Teilhabe im Lesen, Schreiben und Rechnen. [...] leo. operationalisiert dies als Unterschreiten des Alpha-Level 3. leo. differenziert [...] nach Alpha-Level 4, 5 und 6 aus und rechnet das Unterschreiten des Alpha-Level 4 dem funktionalen Analphabetismus zu“ (Grotlüschen, Riekmann & Buddeberg, 2012, S. 18).

Die leo.-Definition des funktionalen Analphabetismus ist somit das Unterschreiten der Textebene, was dem Unterschreiten des Alpha-Level 4 entspricht (Grotlüschen, Riekmann & Buddeberg, 2012, S. 18).

Eine weltweit einheitliche hinreichend operationalisierte Definition existiert jedoch nicht (Grotlüschen & Riekmann, 2012). Die Schwierigkeit einer einheitlichen Definition liegt möglicherweise darin, dass der Sprache und der Sprachfähigkeit linguistische Sprachspezifika zugrunde liegen. Daher erscheint der Weg einer Definition über die gesellschaftliche Teilhabe im Sinne der Vergleichbarkeit sinnvoll. Doch sind nationale Studien verschiedener Länder, in denen das Ausmaß des Funktionalen Analphabetismus angegeben wird, spätestens auf der Ebene der Operationalisierung und der sprachspezifischen Kompetenzmodelle nur schwer oder nur unter Vorbehalt vergleichbar. Daher werden die Größenordnungen des Funktionalen Analphabetismus in Deutschland, Frankreich und Großbritannien im Folgenden dargestellt, ohne einen direkten statistischen Bezug zwischen den Ländern herzustellen. Der Fokus liegt dabei auf der Größenordnung in Deutschland.

(„International Adult Literacy Survey“; vgl. OECD/Statistics Canada 1995) entspricht dies ebenfalls der Kompetenzstufe I des Leseverständnis bei Prosa-Texten.

2.3.3 *Funktionaler Analphabetismus in Deutschland, England und Frankreich*

Im Jahr 2010 wurde deutschlandweit die leo.-Level-One-Studie³¹ durchgeführt. Die leo.-Studie hatte das Ziel, ein Benchmark zur Größenordnung des funktionalen Analphabetismus der deutsch sprechenden Erwachsenen zu definieren. Dabei baut die leo.-Studie auf Vorarbeiten des lea.-Projektes (blogs.epb.uni-hamburg.de/lea) auf. Im lea.-Projekt wurden die Theoriedifferenzierung, die Levelbeschreibungen und die Itementwicklung vorbereitet. Die Studie hatte nicht das Ziel, die Literalität der gesamten Bevölkerung zu erheben, sondern zu ermitteln, wie viele dem niedrigsten Kompetenzbereich – dem sog. Level One – zuzuordnen sind und diesen auszudifferenzieren.

14,5% und somit 7,5 Millionen der erwerbsfähigen erwachsenen Bevölkerung (18-64 Jahre) sind in Deutschland vom funktionalen Analphabetismus betroffen. Wie bereits oben erwähnt bedeutet dies, dass diese Personen die Textebene unterschreiten. Eine Person kann somit zwar einzelne Sätze lesen oder schreiben, nicht jedoch zusammenhängende Texte. Diese Personen sind aufgrund ihrer begrenzten schriftsprachlichen Kompetenzen nicht in der Lage, am gesellschaftlichen Leben in angemessener Form teilzuhaben (Grotlüschen u. a., 2012, S. 20).

Weitere 25,9% (13,3 Millionen) können nur fehlerhaft schreiben. Diese Personen können trotz gebräuchlicher Wörter nur langsam lesen und/oder fehlerhaft schreiben und beherrschen nicht die Rechtschreibung, wie sie bis zum Ende der Grundschule unterrichtet wird. Kumuliert bedeutet dies, dass ca. 40% ihre literalen Kompetenzen noch deutlich verbessern können (Grotlüschen u. a., 2012, S. 20).

Überraschend erscheinen die Ergebnisse zum funktionalen Analphabetismus und der Erwerbstätigkeit: Fast 57% der funktionalen Analphabeten und Analphabetinnen in Deutschland geben an, erwerbstätig zu sein (Grotlüschen, 2012, S. 137).

³¹ blogs.epb.uni-hamburg.de/leo (zuletzt geprüft am 03.11. 2014)

Anteil	Funktionaler Analphabetismus			Summe	Fehlerhaftes Schreiben		Bevölkerung gesamt
Alpha-Level	$\alpha 1$	$\alpha 2$	$\alpha 3$	$\alpha 1 - \alpha 2$	$\alpha 4$	$> \alpha 4$	
Erwerbstätig	54,8%	54,2%	58,0%	56,9%	64,5%	69,5%	66,4%
Arbeitslos	19,6%	21,6%	14,7%	16,7%	8,9%	4,8%	7,6%
Erwerbsunfähig	2,7%	2,3%	2,3%	2,3%	1,5%	0,9%	1,3%
Hausfrau/-mann, Elternzeit	17,45	10,8%	9,4%	10,1%	8,2%	7,9%	8,3%
Rentner/Rentnerin	5,1%	6,3%	6,4%	6,3%	6,2%	3,8%	4,8%
In Ausbildung	0,4%	4,0%	7,9%	6,5%	9,9%	11,6%	10,4%
Sonstiges	0,0%	0,8%	1,4%	1,2%	0,9%	1,4%	1,2%
Summe	100%	100%	100%	100%	100%	100%	100%

Tabelle 3: Beruflicher Status funktionaler Analphabetismus und fehlerhaftes Schreiben (Grotlüschen, 2012, S. 141)

Wie aus der Tabelle 3 hervorgeht, sind fast 55% der Personen auf Alpha-level 1 erwerbstätig, wobei das Alpha-Level 1 bedeutet, dass die zugehörigen Personen die Wortebene beim Lesen und Schreiben nicht erreichen. Von den Personen, die dem Alpha-Level 2 zuzuordnen sind, gehen 54,2% einer Erwerbstätigen Beschäftigung nach. Diese Personen erreichen nicht die Satzebene. Von den Zugehörigen des Alpha-Level 3 sind 58% erwerbstätig, wobei diese Personen kurze Sätze lesen und schreiben können (Grotlüschen & Riekmann, 2011, S. 4). Insgesamt sind somit deutlich mehr funktionale Analphabeten und Analphabetinnen erwerbstätig (56,9%) als arbeitslos (16,7%). Genauere Analysen zeigen, dass 19,7% der funktionalen Analphabeten und Analphabetinnen einer geringfügigen Beschäftigung nachgehen und 86,1% wöchentlich 35 Stunden oder mehr arbeiten. Im Vergleich zur Gesamtbevölkerung sind funktionale Analphabeten und Analphabetinnen überproportional in prekären Beschäftigungsverhältnissen. Allerdings zeichnen sich die prekären Beschäftigungsverhältnisse nicht durch die Befristung von Arbeitsverträgen, sondern in der geringen Entlohnung ab (Grotlüschen, 2012, S. 163).

In England wird zwischen drei Entry Level unterschieden, wobei diese zusammen den IALS Level One abbilden (Grotlüschen & Riekmann, 2012, S. 21). So wird in England von 14,9% funktionaler Analphabeten und Analphabetinnen zwischen 16 und 65 ausgegangen (Department for Business, Innovation and Skills, 2011, S. 5).

In Frankreich wurde mit dem „Information et Vie Quotidienne“ 2004-2005 eine Anzahl von 3,1 Millionen funktionalen Analphabeten und Analphabetinnen erhoben, was neun Prozent der erwachsenen Bevölkerung entspricht. Dabei sind allerdings ausschließlich diejenigen befragt worden, die in Frankreich die Schule besucht haben (ANLCI - French National Agency for the Fight against Illiteracy, 2009).

In Frankreich ist der Term "funktionaler Analphabetismus" allerdings nicht gebräuchlich. In Frankreich wird zwischen drei Formen mangelnder Literalität unterschieden: Personen, die zur Schule gegangen sind, aber Lesen und Schreiben nicht weiter praktiziert haben, Personen, die nie lesen und schreiben gelernt haben sowie Personen, die nicht in Frankreich literarisiert sind und somit nicht in der Lage waren, ihre Lese- und Schreibkompetenzen auszudrücken.

Für die Literalität³² gibt die Studie IALS für Deutschland eine Anzahl von 14,4%, in England von 21,8% auf dem Level One für Literalität aus. Frankreich hatte sich aus der Erhebung zurückgezogen nachdem sich abzeichnete, dass der Wert nahe der Vierzigprozentmarke liegen könnte (Grotlüschen u. a., 2012, S. 23).

Neue Ergebnisse liefert die PIAAC-Studie. In dieser wird zwar nicht die Literalität hinsichtlich funktionalem Analphabetismus ausgewiesen, doch ist die Angabe des Leseverständnisses im Vergleich zu den OECD-Ländern einen Hinweis, wie stark funktionaler Analphabetismus im Vergleich beispielsweise zu Deutschland ausgeprägt sein mag: Frankreich rangiert signifikant deutlich im Bereich Lesekompetenz unter dem OECD-Durchschnitt mit einem Mittelwert von 262 Punkten, Deutschland liegt ebenfalls signifikant unter dem Durchschnitt, weist aber einen höheren Mittelwert von 270 Punkten auf. 18% der deutschen Bevölkerung kommen über die niedrigste Kompetenzstufe von PIAAC nicht hinaus. Spitzenreiter ist Japan mit 296 und Schlusslicht ist Italien mit 250 Punkten (Rammstedt u. a., 2013, S. 13).

Ging man vor der leo.-Studie noch von vier Millionen funktionalen Analphabeten und Analphabetinnen aus, zeigen die leo.- und die PIAAC-Studien doch ein erheblich größeres Ausmaß des funktionalen Analphabetismus. Aller-

³² Im englischen Sprachraum wird von *prose literacy* gesprochen. Unter Prosa-Literalität ist die Literalität zu verstehen, die im Alltag genutzt wird (beispielsweise das Verstehen von Nachrichten). In diesem Kontext ist die Prosa-Literalität gleichbedeutend mit dem Begriff der Literalität im deutschen Sprachraum. Daher wird hier auf den Begriff der Literalität zurückgegriffen.

dings ist anzumerken, dass die Zahl vier Millionen eine Schätzung war, in der nur deutsche Muttersprachler und Muttersprachlerinnen berücksichtigt wurden. In der leo.-Studie wurden nicht nur die in Deutschland literalisierten Personen, sondern auch Personen mit Migrationshintergrund einbezogen. Voraussetzung für die Teilnahme war eine vorgelagerte mündliche Überprüfung, ob die Deutschkenntnisse ausreichen, die Aufgaben verstehen zu können. Somit sind die geschätzte Zahl von vier Millionen und die erhobene Zahl von 7,5 Millionen vor dem Hintergrund der befragungsspezifischen Voraussetzungen nicht als eine Verdopplung oder ein enormer Anstieg des funktionalen Analphabetismus zu betrachten. Zusammenfassend weisen die aktuellen Studien auf einen enormen Alphabetisierungsbedarf hin. Demgegenüber steht die Alphabetisierungspraxis allerdings vor diversen Herausforderungen.

2.3.4 Herausforderungen in der Alphabetisierung in Deutschland

Trotz 7,5 Millionen funktionalen Analphabeten und Analphabetinnen nehmen in Deutschland nur ca. 30.000 Personen an Alphabetisierungskursen teil (Huntemann & Reichart, 2013, S. 30). Ursachen werden sowohl seitens der Betroffenen selbst als auch aus der Sicht von Weiterbildungsanbietern gesehen. Aus der Perspektive von funktionalen Analphabeten und Analphabetinnen werden für die geringe Teilnahme folgende Gründe genannt (Koppel & Wolf, 2014; Steuten, 2013)³³:

- a) geringes Angebot (ca. jede dritte Volkshochschule bietet Alphabetisierungskurse an (Rosenblatt & Lehmann, 2013b, S. 57);
- b) niedriger Bekanntheitsgrad;
- c) Vermeidung, empfundene und/oder tatsächliche Schwächen offen zu legen, beispielsweise auf Grund von Schamgefühlen (vgl. z.B. Döbert u. a., 2000; Egloff, 1997; Füssenich, 2004; Schladebach, 2007);
- d) Subjektiv wird kein Bedarf an Alphabetisierung empfunden.

In den Alphabetisierungskursen selbst sehen sich sowohl Kursleitende als auch Lernende mit weiteren Herausforderungen konfrontiert:

- a) wenig standardisierte Curricula (erst mit der Einbindung der Alpha-Level in das Rahmen-Curriculum der Volkshochschulen im Jahr 2014 wurde eine Grundlage für einheitliche und auf empirisch überprüften Standards basierende Curricula erreicht);
- b) keine einheitlichen Lehr-/Lernmethoden;
- c) keine klare Vorstellung über die Abschlussziele (Rosenblatt & Lehmann, 2013b) (mit dem neuen Rahmencurriculum des Deutschen

³³ Die Herausforderungen in der Alphabetisierungen wurden bereits in Koppel & Wolf (2014) erläutert.

Volkshochschulverbandes ist allerdings die Basis für die Formulierung einheitlicher Ziele geschaffen);

- d) die Nutzung diagnostischer Verfahren ist umstritten (Bonna & Nienkämper, 2011);
- e) Befragungen und Tests stoßen vielfach auf Abwehrreaktionen (Schladebach, 2007), da diese meist eine Defizitorientierung aufweisen.

Die Teilnahme an Alphabetisierungskursen führt somit nur zu einem begrenzten Lernerfolg: Die Lernfortschritte reichen in den meisten Fällen nicht aus, um bei der Mehrzahl der Teilnehmenden ein schriftsprachliches Kompetenzniveau oberhalb des funktionalen Analphabetismus zu erreichen (Rosenblatt & Lehmann, 2013b, S. 72). Aus der Perspektive der Kursleitenden erscheint zudem problematisch, dass wenig erwachsenengerechte Materialien zur Diagnose und Förderung von funktionalen Analphabeten und Analphabetinnen vorhanden sind (Heinemann, 2011). Zudem ist eine individuelle Diagnose des Lernstands meist zeitaufwendig. Auch müsse berücksichtigt werden, dass es um Personen gehe, die häufig Lernbehinderungen aufweisen und in schwierigen sozialen Verhältnissen leben (Rosenblatt & Lehmann, 2013b, S. 57).

Für die Alphabetisierungsarbeit wird daher gefordert, Grundbildungsangebote auf eine tragfähige Basis mit erhöhter Professionalität zu stellen (Frieling & Rustemeyer, 2011, S. 39) sowie Erfolgskriterien und Lernziele zu formulieren (Rosenblatt & Lehmann, 2013b, S. 73).

Als Konsequenzen werden eine Ausweitung und Diversifizierung des bisher relativ schwachen Angebots gefordert (Rosenblatt & Lehmann, 2013a, S. 10). Es sollen differenzierte Angebote geschaffen werden, die sinnvoll die heterogenen Kompetenzstrukturen der funktionalen Analphabeten und Analphabetinnen aufgreifen und berücksichtigen. Auf der Ebene der Kursarbeit sind daher Möglichkeiten bereitzustellen, mit denen zuerst festgestellt werden kann, welche Kompetenzen Kursteilnehmende bereits erworben haben und welche zu fördern sind (Eingangsdiagnostik). Es bedarf daher keiner (summativen) Diagnostik zu Selektionszwecken (vgl. Kap. 2.1), sondern einer Diagnostik zur differenzierten Abbildung der Kompetenzen, die eine Ableitung gezielter Fördermaßnahmen ermöglicht. Eine Möglichkeit, diesen Herausforderungen zu begegnen ist der Einsatz einer computerbasierten Diagnostik. Beim Einsatz computerbasierter Diagnostik ist allerdings immer auch zu berücksichtigen, wie stark die ICT-Literacy ausgeprägt ist und ob die Computerfähigkeit (bzw. ICT-Literacy) einen Einfluss auf das Antwortverhalten hat, um die computerbasierte Diagnostik an die Voraussetzungen der potenziellen Nutzer und Nutzerinnen anpassen zu können.

2.3.5 Informations- und Computertechnologie (ICT)-Literacy

ICT-Literacy wird im Kontext der PISA-Studie folgendermaßen definiert:

„ICT literacy is the interest, attitude, and ability of individuals to appropriately use socio-cultural tools, including digital technology and communication tools, to access, manage, integrate, and evaluate information, construct new knowledge, and communicate with others in order to participate effectively in society“ (Schleicher, 2008, S. 6).

Nach der OECD ist die ICT-Literacy die Voraussetzung bzw. das Werkzeug für effizientes und selbstgesteuertes Lernen mit Informations- und Kommunikationstechnologien (Schäffer, 2006). Diese Definition ist logischerweise an die der Literacy angelegt „the ability to understand and employ printed information in daily activities, at home, at work and in the community – to achieve one’s goals, and to develop one’s knowledge“ (OECD, 2000, S. X [sic!]). Unter Berücksichtigung der Technologieentwicklung wird der Rückbezug auf dieses Konzept in Zukunft nicht ausreichend sein, da Informationen auch auditiv, visuell und wahrscheinlich bald auch haptisch und olfaktorisch dargeboten werden (vgl. auch zur Kritik des ICT-Literacy-Konzepts Schäffer, 2006; Sting, 2005).

Aus länderübergreifenden Studien geht hervor, dass es zwischen den beteiligten Ländern große Unterschiede bezüglich der Computer- und Internetnutzung gibt. So besteht ein signifikanter Zusammenhang zwischen der Höhe des Einkommens, dem formalen Bildungsstand und dem ausgeübten Beruf auf der einen sowie den Literacy-Skills und dem Zugang zum Computer auf der anderen Seite (OECD, 2005, S. 192). Dieses Phänomen wird auch als „digital divide“ beschrieben (vgl. z. B. Montagnier, 2011). Allerdings ist ein zunehmender Anstieg hin zu einer „Vollausstattung“ aller Haushalte mit Computern und Internetzugang in Deutschland zu verzeichnen. Waren im Jahr 2006 noch 64% der Haushalte mit Computern ausgestattet, so waren es 2012 bereits 81% (Statistisches Bundesamt, 2012, S. 27). Der mindestens gelegentliche Internetzugang stieg von 59,5% im Jahr 2006 auf 79,1% im Jahr 2014. Die tägliche Internetnutzung liegt 2014 bei 58,3% (ARD/ZDF-Medienkommission, 2014).

Vom Statistischen Bundesamt werden in regelmäßigen Abständen Erhebungen zu Privaten Haushalten herausgegeben und in Fachserien veröffentlicht. In der Fachserie 15 wird die Nutzung von Informations- und Kommunikationstechnologien dargestellt (Statistisches Bundesamt, 2012). Demnach sind in Deutschland 81% der Haushalte mit Computern ausgestattet, 79% besitzen einen

Internetzugang. In der folgenden Tabelle sind Ergebnisse, aufgeschlüsselt nach Bildungsstand³⁴, aufgeführt:

Nutzung Bildungs- stand	Nutzung Handy/ Smart- phone	Compu- ternutzung jeden/fast jeden Tag	Internetnut- zung inner- halb der letzten drei Monate	Senden/ Empfan- gen von Emails	Mitteilungen in soziale Netzwor- ken, Foren, Blogs oder Chaträumen einstellen
niedrig	76	77	61	88	59
mittel	86	80	77	92	38
hoch	90	88	87	96	34

Tabelle 4: Ausstattung der Haushalte mit Computern 2012 (Statistisches Bundesamt 2012, S.17)

Fast jeden Tag den Computer nutzen 77% mit einem niedrigen Bildungsstand, mit einem mittleren Bildungsstand 80% und mit einem hohen Bildungsstand 88% (Statistisches Bundesamt, 2012, S. 12); 76% mit einem niedrigen, 86% mit einem mittleren und 90% mit einem hohen Bildungsabschluss nutzen ein Handy oder Smartphone (Statistisches Bundesamt, 2012, S. 11). 96% der Personen mit hohem und 88% der Personen mit niedrigem Bildungsabschluss senden und empfangen Emails. Ein umgekehrtes Verhältnis zeigt sich bei den Kategorien „Mitteilungen in soziale Netzwerken, Foren, Blogs oder Chaträumen einstellen“ (34 der Personen mit hohem und 59% der Personen mit niedrigem Bildungsabschluss) und „Telefonieren/Videotelefonate“ (33% und 29%) (Statistisches Bundesamt, 2012, S. 17). Ebenfalls deutliche Unterschiede zeigen sich in den Kategorien „Spiele, Bilder, Filme oder Musik (ab-)spielen/herunterladen“ (60%, 44% und 48%), „Netzwerkspiele mit anderen Personen spielen“ (40%, 26%, 15%) sowie das „Hochladen eigener erstellter Texte, Fotos, Videos, Musik usw. auf Webseiten“ (36%, 27%, 28%) (Statistisches Bundesamt, 2012, S. 18).

Bezüglich der Computerkenntnisse steigt die Diskrepanz zwischen dem Bildungsstand: So können beispielsweise 66% mit einem niedrigen Bildungsabschluss „Kopieren und Einfügen“ von Informationen in ein Dokument, wohingegen 85% mit hohem Bildungsabschluss diese Kenntnis besitzen und anwenden

³⁴ Der in den Tabellen aufgeführte Bildungsstand basiert auf den Bildungsstufen nach ISCED (International Standard Classification of Education), der internationalen Standardklassifikation des Bildungswesens. Diese werden auch von der OECD genutzt. Demnach sind dem niedrigen Bildungsstand die Ausbildungsstufen bis zum Abschluss der Hauptschule, der Realschule, des Gymnasiums (Klassen 5-10), Berufsaufbauschule sowie Berufsvorbereitungsjahr zugeordnet. Dem mittleren Bildungsabschluss sind, Gymnasium, Oberschulen, Duale Berufsausbildung sowie Berufsfachschule zugeordnet. Der Kategorie des hohen Bildungsabschlusses gehören Fachhochschule, Universität, Fachschule, Fachakademie, Schulen des Gesundheitswesens sowie Promotion und Habilitation an.

können; „Komprimieren von Daten“ können 31% mit niedrigem und 52% mit hohem Bildungsabschluss (Statistisches Bundesamt, 2012, S. 13).

Für die vorliegende Arbeit relevant sind insbesondere die Kennzahlen der Personen mit niedrigem Bildungsabschluss. Die Zahlen zur Informations- und Kommunikations-Technologie- (IKT)-Nutzung und den IKT-Kenntnissen sind erste Hinweise darauf, wie funktionale Analphabeten und Analphabetinnen diesbezüglich zu charakterisieren sind. So ist festzuhalten, dass Personen mit niedrigem Bildungsabschluss tendenziell weniger erfahren im Umgang mit Computern im Vergleich zu Personen mit hohem Bildungsabschluss sind. Umgekehrt verhält es sich allerdings im Bereich der Internetaktivitäten in den Kategorien „Mitteilungen in soziale Netzwerke, Foren, Blogs oder Chaträumen einstellen“, „Telefonieren/Videotelefonate, Spiele, Bilder, Filme oder Musik (ab-)spielen/herunterladen“, „Netzwerkspiele mit anderen Personen spielen“ sowie das „Hochladen eigener erstellter Texte, Fotos, Videos, Musik usw. auf Webseiten“. Diese werden häufiger von Personen mit niedrigem Bildungsabschluss als von Personen mit einem hohen Bildungsabschluss genutzt.

Diesbezüglich ist anzumerken, dass die Personen mit niedrigem Bildungsstand zwar die funktionalen Analphabeten und Analphabetinnen einschließen, diese aber nur einen kleinen Teil dieser Gruppe ausmachen. Daher ist von den eben vorgestellten Daten der Gruppe mit niedrigem Bildungsstand nicht auf die Gruppe der funktionalen Analphabeten und Analphabetinnen zu schließen (im empirischen Teil dieser Arbeit werden erste Erhebungen zur Computernutzung und -kompetenz von funktionalen Analphabeten und Analphabetinnen vorgestellt). Weitere Informationen zur ICT-Literacy von Personen im Grundbildungsbereich liefert die aktuelle PIAAC-Studie.

In PIAAC 2012 wurde auch die technologiebasierte Problemlösekompetenz erhoben³⁵, die definiert wurde als „Kompetenz, digitale Technologien, Kommunikationshilfen und Netzwerke erfolgreich für die Suche, Vermittlung und Interpretation von Informationen zu nutzen“ (Rammstedt u. a., 2013, S. 4). 44,9% der Bevölkerung befindet sich hinsichtlich der technologiebasierten Problemlösekompetenz nur auf der Stufe I oder niedriger (Zabal u. a., 2013, S. 70), weitere 17,7% hatten entweder keine Computererfahrung, bestanden die IT-(Vor-) Übung des Tests nicht oder verweigerten sich, den Test computerbasiert durchzuführen. Im OECD-Durchschnitt lagen 41,7% auf oder unter Stufe I. Personen ohne Computererfahrung erreichten durchschnittlich nur deutlich niedrigere Werte in den Bereichen Lesekompetenz (227 vs. 276 Punkte) und alltagsmathematische Kompetenzen (213 vs. 282 Punkte) als Personen, bei denen die Kompetenzmessung computergestützt durchgeführt werden konnte (Zabal u. a., 2013,

³⁵ Die folgenden Überlegungen zu den PIAAC-Ergebnissen werden in Wolf und Koppel (2014) veröffentlicht.

S. 68). Dies lässt den Umkehrschluss zu, dass Personen mit niedriger Lesekompetenz auch weniger Computererfahrung haben und offensichtlich ein gewisser Zusammenhang zwischen der Lesekompetenz, der alltagsmathematischen Kompetenzen und der Computererfahrung besteht, was auch im folgenden Zitat zum Ausdruck kommt:

„Damit neue Technologien kompetent und zweckdienlich zur Lösung von alltäglichen Problemen, zum Beispiel zur Informationsbeschaffung, eingesetzt werden können, sind nicht nur die grundlegenden technische Handhabung von Hard- und Software von Bedeutung, sondern insbesondere auch kognitive Fähigkeiten, wie Lese-, mathematische und Problemlösekompetenz wichtig (International ICT Literacy Panel, 2007)“ (Zabal u. a., 2013, S. 60).

Zusammenfassend lassen die Daten die Schlussfolgerung zu, dass funktionale Analphabeten und Analphabetinnen eine tendenziell niedrige ICT-Literacy vorweisen und diese Voraussetzungen in der Entwicklung und Gestaltung von computerbasierter Diagnostik zu berücksichtigen sind.

2.3.6 *Computerbasierte Lern- und Diagnoseinstrumente im Grundbildungsbereich Erwachsener*

Im Grundbildungsbereich Erwachsener sind bisher wenige Instrumente zur Literalitätsmessung und -förderung vorhanden. Insbesondere fehlen Instrumente, die eine kompetenzorientierte Diagnostik anbieten und somit auch das Potenzial für eine förderdiagnostische Tätigkeit besitzen. In Deutschland werden insbesondere zwei computerbasierte Instrumente für den Grundbildungsbereich eingesetzt: das Lernportal „ich-will-lernen.de“³⁶ und das Lernadventure „Winterfest“³⁷. Beide Instrumente wurden vom BMBF gefördert. „ich-will-lernen.de“ ist Deutschlands größtes offenes Lernportal mit mehr als 31.000 kostenlosen Übungen zur Alphabetisierung und Grundbildung. Das Portal eignet sich für Einsteigende als auch Personen mit Vorkenntnissen und kann anonym genutzt werden. Seit dem Jahr 2004 wurden mehr als 389.000 Passwörter vergeben (BMBF, 2014). Das Lernadventure Winterfest ist ein Lernspiel, ebenfalls für Jugendliche und Erwachsene. Aufgaben und Übungen sind in Geschichten eingebettet; Minispiele und Rätsel sollen die Teilnehmenden motivieren, ihre Lese-, Schreib- und Rechenfähigkeiten zu verbessern. Das Spiel kann sowohl privat als auch in der

³⁶ www.ich-will-lernen.de (zuletzt geprüft am 03.11. 2014)

³⁷ www.lernspiel-winterfest.de (zuletzt geprüft am 03.11. 2014)

Kursarbeit eingesetzt werden. Zusätzlich existieren Lehr- und Lernmaterialien, welche ergänzend zum Spiel eingesetzt werden können.

Der Fokus liegt bei diesen Plattformen allerdings nicht auf der Kompetenzmessung, sondern auf dem Lernen von Lesen, Schreiben und Rechnen. Diesen Instrumenten liegt kein validiertes Kompetenzmodell zu Grunde. Somit sind diese für die Diagnostik und insbesondere für die Förderdiagnostik nur bedingt geeignet. Ein Instrument, das förderdiagnostischen Prinzipien genügen soll ist die Online-Testumgebung *otu.lea*. Dessen Entwicklung und Evaluation wird im weiteren Verlauf dieser Arbeit vorgestellt.

Auf internationaler Ebene existieren zahlreiche weitere Online-Plattformen zum Lernen von Lesen, Schreiben und Rechnen. An dieser Stelle soll lediglich einige wenige dieser Programme kurz vorgestellt werden. Die Lernplattform *GCFLearnFree.org*³⁸ bietet Lernmaterialien und Aufgaben im Bereich Technologie (Computerfähigkeiten wie beispielsweise Nutzung von Word und Excel), Lesen (sowohl für Muttersprachler und Muttersprachlerinnen als auch für Personen, die Englisch als Zweitsprache lernen) und Mathematik. Insgesamt stellt die *GCFLearnFree.org* über 85 Übungsthemen und 750 verschiedene Übungseinheiten kostenlos zur Verfügung, die bisher von mehreren Millionen Personen in über 200 Ländern genutzt wird. Allerdings liegt auch hier kein validiertes Kompetenzmodell zugrunde.

Computerbasierte Diagnoseinstrumente für den Grundbildungsbereich Erwachsener in England sind z. B. der *Target Skills: Initial Assessment*³⁹ und der *Skills for Life Survey*⁴⁰. Beide sind jedoch nicht frei zugänglich und werden daher im Folgenden nur kurz beschrieben.

Target Skills bietet eine computerbasierte Diagnostik und Lernmaterialien in mehreren Dimensionen an: Basic Skills Screener (ein Test bezüglich grundlegender Fertigkeiten in Literalität und mathematischen Grundfertigkeiten), Entry Level sowie Level 1 für Literacy und Numeracy (Lernmaterialien für Erwachsene, die sich auf das Kerncurriculum in den Entry-Level beziehen). Die Materialien sind frei zugänglich, müssen aber käuflich erworben werden. Zum *Target Skills: Initial Assessment* berichteten Brooks u.a. im Report vom NRDC, dass dieser nicht für formative Zwecke geeignet und zudem aufgrund eines möglichen Einflusses der ICT-Literacy nicht ausreichend valide sei (Brooks u. a., 2005, S. 105). Beim *Skills for Life Survey* handelt es sich um eine nationale Studie in England, in der Literalitäts-, Mathe- und Computerkompetenzen erhoben werden (vgl. auch Kap. 2.3.3). Das Instrument ist nicht öffentlich und wird von

³⁸ www.gcflearnfree.org (zuletzt geprüft am 03.11. 2014)

³⁹ www.targetskills.net (zuletzt geprüft am 03.11. 2014)

⁴⁰ <https://www.gov.uk/government/organisations/departement-for-business-innovation-skills> (zuletzt geprüft am 03.11. 2014)

Brooks u. a. hinsichtlich mangelndem *gender* und *cultural mainstreaming*, einem möglichen Einfluss der ICT-Literacy sowie mangelnder Authentizität kritisch bewertet (Brooks u. a., 2005, S. 118). Der Test ist in seinen jeweiligen Abschnitten adaptiv.

Zusammenfassend werden inzwischen Computer in internationalen Vergleichsstudien zur Messung der Literalität (z. B. PIAAC, Skills for Life, IALS) eingesetzt. Frei zugängliche computerbasierte Diagnoseinstrumente liefern jedoch nach dem gegenwärtigen Erkenntnisstand keine Diagnostik, die eine differenzierte und kompetenztheoretisch überprüfte Rückmeldung ermöglichen.

Was insbesondere aus dem Report von Brooks u. a. (2005) hervorgeht, ist der mögliche Einfluss der ICT-Literacy auf das Testergebnis. Um valide Ergebnisse zu erzielen, muss der Einfluss der Computerkompetenz weitergehend ausgeschlossen werden können. Computerbasierte Diagnoseinstrumente sind daher so zu gestalten, dass die Nutzer und Nutzerinnen nicht kognitiv allein aufgrund der Aufgabendarstellung überfordert werden. Hinweise auf mögliche Einflussfaktoren liefern u.a. die Kognitionspsychologie - insbesondere die CLT sowie die CTML (Kap. 2.4) -, die Äquivalenzproblematik (Kap. 2.5) sowie die Usability-Forschung (Kap. 2.6). In den folgenden Kapiteln werden diese möglichen Einflussfaktoren näher erläutert.

2.4 Cognitive Load Theory und Cognitive Theory of Multimedia Learning

Die CLT und die CTML liefern Modelle über die Wirkungsweise von Auslastungsfaktoren auf das Arbeitsgedächtnisses. Aus beiden Theorien lassen sich Überlegungen ableiten, wie computerbasierte Diagnoseinstrumente gestaltet werden können, um den Einfluss der ICT-Literacy auf das Diagnoseergebnis zu minimieren bzw. zu verhindern.

2.4.1 Grundidee und Annahmen der Cognitive Load Theory

Die Grundidee der CLT besteht in der Erforschung der kognitiven Belastung beim Lernen. Sie hat zum Ziel, psychologische Phänomene zu erklären und Hinweise für die Informationsdarstellung zu geben, um die Informationsverarbeitung zu optimieren (Moreno & Park, 2010, S. 9). Lernende sollen darin unterstützt werden, die verfügbaren kognitiven Ressourcen nicht für die Erschließung irrelevanter Informationen des Aufgabenmaterials, sondern für die Verarbeitung von Informationen und das tiefere Verständnis zu nutzen. Wie auch der im Vorfeld bestehenden Theorie des „mental Load“ (Moray, 1979) liegt der CLT die Annahme zugrunde, dass das Arbeitsgedächtnis begrenzt und das Langzeitgedächtnis unbegrenzt sind. In neueren Begriffsbestimmungen repräsentiert der *Mental Load* die Anforderungen an die kognitiven Ressourcen, die aus der Interaktion sowohl zwischen Aufgaben- als auch Subjektcharakteristiken bei einer

Instruktion resultieren (Cook, Zheng & Blaz, 2009, S. 38). In der Theorie von Moray wurden allerdings die psychologischen Effekte auf die Auslastung des Arbeitsgedächtnisses (Einstellungen, Erwartungen, individuelle Ziele) nicht berücksichtigt (Moreno & Park, 2010, S. 10). In der CLT werden hingegen personale bedingt Voraussetzungen einbezogen. Erste Skizzierungen der Theorie und die Verwendung des Begriffs Cognitive Load finden sich in Artikeln von Sweller (Sweller, 1988, 1989), wobei sie schließlich von Chandler und Sweller (1991) konkretisiert wurden. Im Fokus der CLT (CLT) steht das Arbeitsgedächtnis, welches dem Modell von Baddeley (1976) zufolge neben dem Langzeitgedächtnis einen Bestandteil des menschlichen Gedächtnisses darstellt. Das Arbeitsgedächtnis wird benötigt, um aktuelle Prozesse zu bearbeiten, um neues Wissen entweder an bereits Vorhandenes anzuknüpfen oder um neue Schemata zu entwickeln. Es nimmt Informationen über den visuellen/piktographischen oder den auditiven/verbalen Kanal auf. Die Annahmen der CLT werden besonders bei der Gestaltung von Instruktionen im Lernkontext berücksichtigt (Paas, Tuovinen, Tabbers & van Gerven, 2003, S. 63). Ein wesentliches Merkmal ist, dass die Kapazität des Arbeitsgedächtnisses und die Aufnahmefähigkeit begrenzt sind. Als Auslastungsfaktoren des Arbeitsgedächtnisses beim Lernen gelten u. a. das Aufgabenformat, die Aufgabenkomplexität und der Einsatz von Multimedia. Die Auslastung des Arbeitsgedächtnisses ist dabei die subjektiv wahrgenommene Aufgabenschwierigkeit, beeinflusst durch Motivation, Fähigkeit, Erwartung, Training, Zeit, Stress, Ermüdung und Umstände (Kantowitz, 1987, S. 97). Der CLT zufolge lässt sich das Arbeitsgedächtnis in drei Bereiche unterteilen: *Intrinsic Load*, *Extraneous Load* und *Germane Load* (Sweller, 2010, S. 40). Zusätzlich zu den drei Bereichen besteht freie Kapazität (vgl. Abbildung 2).

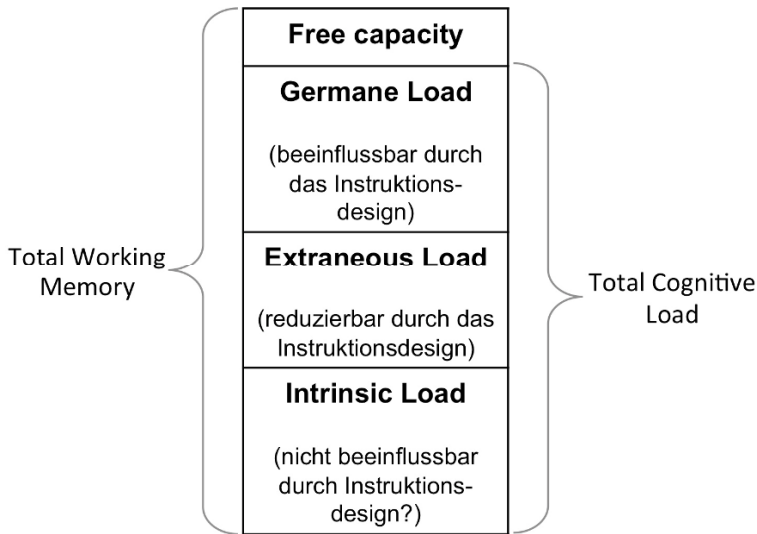


Abbildung 2: Arbeitsgedächtnis und Cognitive Load (Moreno & Park, 2010, S. 18)

Der *Intrinsic Load* ist der Anteil kognitiver Belastung, der in dem intellektuellen Anspruch des Lernmaterials begründet ist. Auslastungsfaktor ist die Element-Aktivität, die durch die Anzahl der gleichzeitig im Arbeitsgedächtnis zu verarbeitenden einzelnen Wissensinhalte entsteht. Lerninhalte, die lediglich den Abruf einzelner isolierter Wissensbestandteile erfordern, wie beispielsweise das Vokabeln lernen, beanspruchen den *Intrinsic Load* nur wenig im Vergleich zu Wissensinhalten, die in einem komplexen Zusammenhang stehen und für das Verständnis des Lerninhalts gleichzeitig im Arbeitsgedächtnis repräsentiert sein müssen (Sweller, 2005). Der *Extraneous Load* entsteht durch den überflüssigen Anteil an Beanspruchung durch eine nicht optimierte Gestaltung des Lernmaterials. Ist das Lernmaterial beispielsweise unübersichtlich gestaltet, müssen kognitive Ressourcen beansprucht werden, um die relevanten Informationen zu identifizieren. Dies würde eine vergleichsweise hohe Auslastung des *Extraneous Load* zur Folge haben. Bei dem *Germane Cognitive Load* handelt es sich um den Anteil kognitiver Beanspruchung, welcher für das Lernen wichtig ist und zu dessen Gunsten es gilt, den *Extraneous Load* möglichst gering zu halten. Der *Germane Load* wird während der tieferen Verarbeitung von Informationen sowie der Bildung neuer oder der Anpassung bereits bestehender Schemata (und somit zum Lernen) beansprucht. Allerdings besteht eine Beanspruchung des *Germane Load* nur, wenn auch der *Intrinsic Load* eine Auslastung erfährt. Ist der *Intrinsic Load*

niedrig und der *Extraneous Load* hoch, ist dennoch Kapazität für den *Germane Load* vorhanden; jedoch kann kein großer Lerneffekt stattfinden, da das Arbeitsgedächtnis vorwiegend mit externen Faktoren (Darstellung der Inhalte) und nicht mit den Inhalten selbst ausgelastet ist. Folglich steigt die Kapazität des *Germane Load*, wenn auch der *Intrinsic Load* steigt (Sweller, 2010, S. 44). Effekte des *Extraneous Load* können daher nur beobachtet werden, wenn auch der *Intrinsic Load* beansprucht wird (Sweller, 2005, S. 28). Der *Intrinsic* und der *Extraneous Load* sind additiv, d. h. wenn die Auslastungsgrenze des Arbeitsgedächtnisses erreicht ist, ist das tiefere Verständnis und somit das Lernen gefährdet. Im Gegensatz zum *Intrinsic Load* sind *Extraneous* und *Germane Load* von dem Aufgabendesign beeinflussbar (Sweller, 2010, S. 44).

Schlussfolgernd hat die Gestaltung der Aufgabe einen direkten Einfluss auf die Kapazitätsauslastung des Arbeitsgedächtnisses und insbesondere auf den *Extraneous Load*. Ist die Auslastung durch das Aufgabenformat hoch, ist weniger Arbeitsspeicher für die Verarbeitung der Informationen vorhanden.

Studien zeigen allerdings, dass die Beanspruchung des Cognitive Load in einem Zusammenhang mit dem Vorwissen, räumlichen Denk- und Vorstellungsvermögen, Selbstregulation sowie Motivation im Allgemeinen stehen (Low & Jin, 2009). Erklären lassen sich die Zusammenhänge zwischen Cognitive Load und dem Vorwissen, dem räumlichen Denk- sowie dem Vorstellungsvermögen mit den Grundprozessen des Lernens: Lernen geschieht, wenn neue Schemata entwickelt werden (Akkommodation) oder neue Elemente in bereits bestehende Schemata integriert werden (Assimilation) (vgl. z. B. Graf, Zimbardo & Gerrig, 2007, S. 66). Die Schemata werden im Langzeitgedächtnis gespeichert und bei Bedarf im Arbeitsgedächtnis präsentiert. Dabei werden sie als einzelne Elemente (*Chunks*) behandelt. Können die Schemata zunehmend automatisiert abgerufen werden, „verschwindet“ zunehmend die Grenze des Arbeitsgedächtnisses, da die Informationen direkt in das Langzeitgedächtnis in bereits vorhandene Schemata integriert werden können (Kirschner, Kester & Corbalan, 2011, S. 1). Das Vorwissen kann wiederum dazu beitragen, die Auslastung des *Extraneous Load* zu reduzieren, indem einzelne Elemente des Informationsmaterials zu größeren Einheiten zusammengefasst werden. Dieser Prozess wird als *chunking* bezeichnet (Miller, 1956). Andersherum kann einer starken Beanspruchung des Arbeitsgedächtnisses entgegengewirkt werden, indem Chunks in kleinere Einheiten gesplittet werden.

Kritisiert wird an der CLT, dass die Motivation, das Vorwissen und auch die Anstrengungsbereitschaft nicht explizit berücksichtigt sind. Auch wenn unbestritten ist, dass die Motivation einen Einfluss auf die Auslastung des Arbeitsgedächtnisses und somit auf den Lernerfolg hat und die Selbstregulationsstrategien und die Motivation das Lernen positiv beeinflussen können (Deci & Ryan,

1993), sind die Erkenntnisse über den Zusammenhang zwischen Motivation und Cognitive Load rar (Low & Jin, 2009, S. 154). Es wäre zu untersuchen, in wie weit eine hohe bzw. niedrige Motivation die Beanspruchung des Arbeitsgedächtnis beeinflussen. Schlussfolgernd gilt es einerseits, die Faktoren zu berücksichtigen und andererseits den Einfluss der Motivation auf die Auslastung des Arbeitsgedächtnisses stärker zu erforschen.

2.4.2 *Cognitive Theory of Multimedia Learning*

Die CTML differenziert den Ansatz der CLT aus und bezieht diesen speziell auf das Lernen mit Multimedia. Der Einsatz von Multimedia ermöglicht die Informationsweitergabe sowohl über den auditiven/verbalen als auch den visuellen/piktographischen Kanal. Einerseits kann diese duale Beanspruchung die Informationsaufnahme positiv beeinflussen bzw. vereinfachen. Andererseits können durch verschiedene Reize das Interesse, die Aufmerksamkeit und die Lernmotivation gesteigert werden (Moreno, 2005, S. 508). Es ist daher davon auszugehen, dass eine Anreicherung von Aufgabenformaten durch Multimedia die Informationsaufnahme (z. B. Instruktion und Fragestellung in einem E-Assessment) erleichtert. Folglich hat dies einen reduzierenden Effekt auf die Auslastung des Cognitive Load, so dass mehr Gedächtniskapazität vorhanden ist, um die Aufgabe zu bearbeiten. Zudem kann die Einbindung von Multimedia die Aufmerksamkeit und Motivation steigern.

Mayers Theorie der kognitiven Belastung beim Lernen mit Multimedia liegen drei Annahmen zu Grunde (Mayer, 2005, S. 33 ff):

1) *Dual Channel Assumption*: Das Arbeitsgedächtnis beinhaltet zwei unterschiedliche Systeme bzw. Kanäle für die Enkodierung und Verarbeitung von visuellen/piktographischen und auditiven/verbalen Inhalten.

2) *Limited Capacity Assumption*: Die Ressourcen für die Repräsentation in den jeweiligen Kanälen (visuell/piktographischen und auditiv/verbal) sind begrenzt. Wenn die Prozesse die Ressourcen überschreiten, entsteht der sog. *Cognitive Overload*. Die Begrenzung der Kanäle führt dazu, dass Entscheidungen darüber getroffen werden müssen, welchen Informationen Aufmerksamkeit geschenkt wird und zu welchem Grad Verknüpfungen zwischen den Informationen bzw. zwischen den Informationen und bereits vorhandenem Wissen hergestellt werden sollen. Hierbei können metakognitive Strategien hilfreich sein.

3) *Active Processing Assumption*: Bedeutungsvolles Lernen findet statt, wenn die Personen aktiv und simultan die Informationen sowohl vom visuellen als auch vom auditiven Kanal verarbeiten. Beim Lernen mit Multimedia laufen dabei fünf Prozesse ab (Mayer, 2005, S. 38): 1. Auswahl relevanter Wörter für die Repräsentation im verbalen Arbeitsgedächtnis, 2. Auswahl relevanter Bilder für die Repräsentation im visuellen Arbeitsgedächtnis, 3. Organisation der aus-

gewählten Wörter in ein verbales Modell, 4. Organisation der ausgewählten Bilder in einem piktographischen Modell und 5. Integration der verbalen und piktographischen Repräsentationen untereinander und mit dem vorhandenen Vorwissen (vgl. z. B. auch J. R. Anderson, 2013).

Für das Designen von Benutzer-/Benutzerinnenoberflächen sind somit visuelle und auditive Informationen aufeinander abzustimmen, so dass die Aufnahme der Informationen weniger Ressourcen beansprucht als wenn die Informationen in ihrer unterschiedlichen Präsentationsform nicht aufeinander abgestimmt und einzeln verarbeitet werden müssten. So können die Informationen schneller und "einfacher" mit bereits bestehenden Modellen abgeglichen und ggf. angepasst (assimiliert) werden.

Dass eine angemessene Darstellung beim multimedialen Lernen die kognitive Belastung reduzieren können, belegen zahlreiche Studien (Mayer & Anderson, 1991; Mayer & Moreno, 2003; Mayer, Moreno, Boire & Vagge, 1999; Mayer & Sims, 1994). Doch führt wiederum eine für den Lernenden nicht adäquate Präsentation multimedialer Inhalte zu einer Steigerung der kognitiven Belastung und behindert somit das Lernen. Zudem darf der Medienreichtum allerdings nicht zu einer „Verkomplizierung“ (*overcomplication*) führen, noch darf es zu einer zu starken „Vereinfachung“ (*oversimplification*) kommen (Schwabe, 2001, S. 55). Gerade in Bezug auf funktionale Analphabeten ist ein geringer Toleranzbereich zu vermuten. Das bedeutet, die Medienwahl muss möglichst zu deren Lese- und Schreibkompetenz sowie zu deren ICT-Literacy passen.

Aus den Annahmen zur CTML sind als Konsequenzen für die Gestaltung von Multimedia insbesondere zwei Prinzipien zu folgern: 1) das präsentierte Material soll eine kohärente Struktur aufweisen; 2) die dargebotene(n) Informationen solle(n) Orientierungsmöglichkeiten geben, um sich als Rezipient die Struktur erschließen zu können. Diese Prinzipien finden sich auch in Gestaltungsrichtlinien für Usability wieder (vgl. Kap. 2.6.3.2, z. B. „Gesetz der Nähe“).

Kritisiert wird an der CTML, dass die Motivation des Lernens außer Acht gelassen wird, welche eine einflussreiche Rolle spielt (Merriënboer & Sweller, 2005). Eine Reduktion der kognitiven Belastung führt nur zu einem Lerneffekt, wenn die lernende Person motiviert ist zu lernen, die freien Ressourcen zu nutzen und in den *Germane Load* zu investieren. Problematisch an diesem Konzept der Kognitiven Belastung ist zudem, dass von einer kumulativen und statischen Beschaffenheit ausgegangen wird. Beim Lernen – und insbesondere beim komplexen Lernen – sind mehrere Vorgänge und Entscheidungen involviert, welche die darauf folgenden lernrelevanten Aspekte beeinflussen. Und so erwähnen Cook u. a., dass auch der *Cognitive Load* als dynamisches Konstrukt betrachtet werden müsse (Cook u. a., 2009, S. 38). Lösungsansätze liefern beispielsweise

Xie & Salvendy (2000) mit der Differenzierung zwischen *instantaneous Load*, *peak Load*, *accumulated Load*, *average Load* und *overall Load*.

Die CLT und die CTML unterscheiden sich darin, dass bei der CLT die Interaktionen von Informationen und weniger der Einsatz von Multimedia im Vordergrund stehen, wohingegen bei der CTML der Fokus auf mentalen Repräsentationen durch Multimedia liegt. Mayer sieht den Unterschied zudem darin, dass bei der CLT die 5 Prozesse (Auswahl relevanter Wörter/Bilder für die Repräsentation im verbalen/visuellen Arbeitsgedächtnis, Organisation der ausgewählten Wörter/Bilder in ein verbales/piktographischen Modell, Integration der verbalen und piktographischen Repräsentationen untereinander und mit dem vorhandenen Vorwissen) nicht berücksichtigt werden (Brünken, Plass & Leutner, 2004).

Bisher haben die Ansätze der CLT und CLML nur im Kontext von Lernen und Lehren mit Multimedia stattgefunden, nicht aber im Bereich der Kompetenzdiagnostik im Allgemeinen und des E-Assessments in Verbindung mit der probabilistischen Testtheorie (IRT – Item Response Theory)⁴¹ im Besonderen.

Es ist davon auszugehen, dass eine hohe Usability (vgl. Kap. 2.6) eine geringe Auslastung des *Extraneous Load* zur Folge hat. Anders herum formuliert: Eine schlechte Usability würde den Nutzer/die Nutzerin kognitiv stark beanspruchen – beispielsweise kann die Orientierung auf einer Website schwierig sein, wenn diese unübersichtlich gestaltet ist (vgl. Tabelle 5).

⁴¹ Die Aufgaben der Online-Testumgebung wurden auf der Basis der IRT entwickelt. Eine genauere Erläuterung hierzu erfolgt in Kap. 4, in dem die Entwicklung der Online-Testumgebung und der Projektkontext erläutert werden.

Usability Auslastung	ausgeprägte Usability ohne Einsatz von Multimedia	ausgeprägte Usability mit Einsatz von Multimedia	Usability niedrig (mit/ohne Einsatz von Multimedia)
Extraneous Load	mittel	niedrig	hoch
Intrinsic Load	hoch	hoch	hoch
Germane Load	hoch	niedrig	niedrig

Tabelle 5: Angenommener Zusammenhang zwischen der Beanspruchung des Cognitive Load und Usability

Die kognitive Beanspruchung würde vermutlich umso stärker ausfallen, je weniger Computererfahrung die Personen haben. Personen mit wenig Computererfahrung haben bisher wenig mentale Modelle bzw. Schemata hinsichtlich Benutzer-/Benutzerinnenoberflächen (z. B. hinsichtlich eines typischen Seitendesigns und/oder Seitenstruktur sowie grundlegender Funktionen) gebildet, die sie abrufen könnten. Daher ist der *Extraneous Load* bei Personen mit wenig Computererfahrung und -kompetenz vermutlich stärker ausgelastet als bei Personen mit mehr Computererfahrung und -kompetenz (Koppel, 2011; Mayer, 2005; Plass, Moreno & Brünken, 2010). Der *Extraneous Load* würde aufgrund der Computerkompetenz und -erfahrung sowie möglicherweise auch aufgrund der Seitengestaltung zu Lasten des *Germane Load* stark ausgelastet sein und durch eine gleichzeitige Repräsentation von komplexen Wissensbestandteilen die Lösung der Aufgabe beeinträchtigen.

2.4.3 Messmöglichkeiten des Cognitive Load

Wie bereits oben beschrieben ist eine zentrale Annahme, dass das Instruktionsdesign und oder die Methoden entweder eine das Lernen unterstützende oder eine dem Lernen hinderliche (durch eine hohe Beanspruchung des *Extraneous Load*) Auslastung zur Folge hat. Eine weitere Annahme besagt, dass der Cognitive Load in Abhängigkeit der Aufgabenkomplexität variiert und davon auch die Beanspruchung des *Intrinsic Load* abhängt (Brünken, Seufert & Paas, 2010, S. 181). Die bisherigen Erläuterungen und Annahmen deuten bereits darauf hin, dass der Cognitive Load nicht direkt beobachtbar ist, sondern nur indirekt gemessen werden kann. In den vergangenen 15 Jahren konnten weder standardisierte Messmethoden noch ein „allgemeingültiges“ Forschungsparadigma entwickelt werden. Die bisherigen Messmethoden werden in analytische, aufgaben- und performanzbasierte Methoden sowie physiologische Techniken und Beurteilungsskalen kategorisiert. In Anlehnung an Brünken u. a. (2010) können diese wiederum in zwei Bereiche eingeteilt werden:

1) Subjektive Methode bzw. Selbsteinschätzung: Personen werden nach ihrer wahrgenommenen Auslastung befragt;

2) Objektive Messmethoden: Es wird die physiologische Performanz gemessen.

Zu 1) Die subjektive Messmethode werden Personen gebeten, ihre wahrgenommene Gedächtnisauslastung auf einer Skala zwischen „gar nicht“ und „sehr viel“ einzuschätzen. Die Skalenmethode basiert auf der Annahme, dass die Personen eine valide Einschätzung bezüglich ihrer Gedächtnisleistung und der zu bewältigenden Aufgabe abgeben können. Die meisten Methoden nutzen eine 7-9-stufige Likert-Skala. Kombiniert werden die Methoden oftmals mit der Einschätzung der Aufgabenschwierigkeit. Bisherige Studien weisen eine hohe Korrelation zwischen den beiden Dimensionen (Selbsteinschätzung der Gedächtnisauslastung sowie wahrgenommene Aufgabenschwierigkeit) auf, so dass auch eindimensionale Untersuchungen durchaus valide Ergebnisse produzieren. Der größte Vorteil dieser Methoden liegt in dem geringen Aufwand der Umsetzung. Allerdings weist diese Methode auch bedeutsame Grenzen auf, denn Personen schätzen im *Nachhinein* die wahrgenommene Gedächtnisleistung ein (Brünen u. a., 2010, S. 183). Es ist nicht ersichtlich, ob bzw. welche Faktoren diese Selbsteinschätzung beeinflussen. Diese Faktoren können sich sowohl durch Rahmenbedingungen (z. B. Lautstärke in dem Untersuchungsraum) als auch durch physiologische Gegebenheiten (z. B. Müdigkeit) generieren. Andererseits kann dieser Effekt ausgeschlossen werden, wenn zu unterschiedlichen Zeitpunkten eines Tests die Selbsteinschätzung wiederholt werden. Eine weitere größere Einschränkung ergibt sich aus der Inhaltsvalidität: Für die Personen, die eine Selbsteinschätzung vornehmen ist meist nicht deutlich zu differenzieren, welcher Bereich des Cognitive Load ausgelastet ist (Brünen u. a., 2010, S. 183). Ein Ansatz für die Messung des *Extraneous Load* kann sein, Aufgaben mit gleicher Aufgabenschwierigkeit aber unterschiedlich dargebotenen Formaten bearbeiten zu lassen.

Zu 2) Es existieren zahlreiche Indikatoren zur objektiven Messung des *Cognitive Load*. Bezüglich des Lernprozesses können diese in „Outcome-Variablen“, „Input-Variablen“ sowie „Prozessorientierte Variablen“ (z. B. Performanz oder Bearbeitungszeit) differenziert werden, wobei der objektivste Indikator der Lernzuwachs ist (Brünen u. a., 2010, S. 183). Messmöglichkeiten bestehen in experimentell kontrollierten Lernsituationen. Fällt der Lernzuwachs unterschiedlich aus, können diese Unterschiede in der unterschiedlichen Auslastung des *Cognitive Load* begründet sein. Grenzen dieser Methode bestehen darin, dass zwar die abhängige Variable Lernzuwachs und die unabhängigen Variablen der experimentellen Bedingungen benannt werden können. Aber es kann nicht eindeutig festgestellt werden, ob tatsächlich der *Cognitive Load* den Effekt auf

den Lernzuwachs auslöst oder ob andere Einflussfaktoren eine Rolle spielen und die Ursachen für diese Varianz sind (z. B. die Motivation). Der Lernzuwachs selbst stellt somit keinen erfüllenden und zureichend erklärenden Indikator dar, dennoch kann der Lernzuwachs in Kombination mit weiteren Messmethoden hinreichende Validität liefern.

Cook u. a. (2009, S. 42) stellen die Messmethoden in einer Matrix dar und bewerten sie hinsichtlich ihrer Subjektivität/Objektivität und der Auslastung, die durch die Methode gemessen werden kann. Dabei ordnen sie ebenfalls die Beurteilungsskalen und analytischen Vorgehen den subjektiven und Performanz sowie physiologische Messungen (beispielsweise der Herzfrequenz oder Gedächtnisaktivität) den objektiven Messmethoden zu. Allerdings legen sie nicht die Differenzierung zwischen *Intrinsic* und *Extraneous Load*, sondern die von Xie & Salvendy (2000) zugrunde (*instantaneous Load*, *peak Load*, *accumulated Load*, *average Load* und *overall Load*, vgl. vorheriges Kap.). Die Wahl der Messmethoden hängt sowohl vom Kontext als auch von der Fragestellung ab. Kontextbedingt ist der Einsatz bestimmter Verfahren aufgrund der Zielgruppe oder der zur Verfügung stehenden Ressourcen ausgeschlossen. Beispielsweise bedeutet der Einsatz von Instrumenten für die Messung der Gedächtnisaktivität einen hohen finanziellen Aufwand und ist für Forschende nicht immer tragbar. Schließlich sei angemerkt, dass es sich bei dem Begriff Cognitive Load um ein Konstrukt handelt, welches immer nur indirekt gemessen werden kann. Eine direkte Beobachtung ist nicht möglich. Dennoch liefern die CLT und CTML wesentliche Hinweise für die Gestaltung von Benutzer-/Benutzerinnenoberflächen und somit auch für die Gestaltung von computerbasierten Instrumenten. Insbesondere die Annahmen über den *Intrinsic* und den *Extraneous Load* weisen darauf hin, dass die Gestaltung der Benutzer-/Benutzerinnenoberflächen von computerbasierten Diagnoseinstrumenten einen Einfluss auf die Performanz und daher auch auf das zu erfassende Konstrukt ausüben kann. Insbesondere bei einer direkten Übertragung von einer papier- in eine computerbasierte Version muss kritisch hinterfragt werden, ob sich durch die Darstellungsform der Items das Testergebnis verändert, worauf im folgenden Kapitel eingegangen wird.

2.5 Exkurs: Äquivalenzproblematik

Die Äquivalenzproblematik behandelt die Schwierigkeit eines Transfers von einer papier- auf eine computerbasierte Testung.

„There is one methodological issue that should be considered from a technological point of view, however, and this is validity. Different validity issues

may arise when TBA [Technology Based Assessment] is applied to replace traditional paper-based assessment and when skills related to the digital world assessed“ (Csapó u. a., 2012, S. 145).

Zu Beginn der Einsatzphase von computerbasierter Diagnostik wurden oftmals bereits bestehende papierbasierte Instrumente in eine technologiebasierte Version transferiert. Doch wenn der Testmodus geändert wird – z. B. von papierbasiert (PP für *Paper Pencil*) zu computerbasiert – kann sich unter Umständen die Natur des erfassten Konstrukts verändern, so dass zwischen den beiden Versionen keine Äquivalenz mehr besteht.

Eine Studie von Pomplun, Ritchie & Custer (2006) deutet auf die Äquivalenzproblematik hin: Trotz eines 1:1-Transfers von einem papier- zu einem computerbasierten Erhebungsinstrument wurden unterschiedliche Effekte erzeugt. In der Untersuchung wurden Faktoren extrahiert und untersucht, welche auf unterschiedliche Effekte einer technologiebasierten respektive papierbasierten Befragung hinweisen. Relevante Faktoren beziehen sich auf Probanden-/ Probandinnen- und Itemcharakteristikvariablen. Die Studie wurde bei Grundschüler und -schülerinnen durchgeführt, welche zu dem Zeitpunkt der Testdurchführung niedrige Lese- und Schreibfähigkeiten hatten. Die Ergebnisse hinsichtlich des Itemformats zeigen, dass Probanden und Probandinnen in der papierbasierten Befragung einen höheren Wert erzielten, als im computerbasierten Assessment (Pomplun u. a., 2006, S. 135). Zudem fielen die Effekte um so größer aus, je weniger die Person mit dem Computer gearbeitet hat (Pomplun u. a., 2006, S. 129). Zusammenfassend zeigt diese Studie, dass Differenzen im Antwortverhalten aus den Unterschieden der Erhebungsformen (PP vs. computerbasiert) resultieren und in der ICT-Literacy begründet sein können.

Auch die Studie von Bennett (2008) belegt, dass Testergebnisse im computerbasierten Assessment niedriger als im Vergleich zu den Ergebnissen der papierbasierten Erhebung ausfielen und das Maß der Vertrautheit im Umgang mit Computern eine Rolle spielt. Weiterhin wurde deutlich, dass der Unterschied zwischen PP-Erhebungen und computerbasierten Assessments in Abhängigkeit zum Itemformat steht: Der Effekt auf das Testergebnis war größer, wenn Befragungsteilnehmende Sätze konstruieren mussten, als der Effekt von Multiple-Choice Items im computerbasiert-PP-Vergleich (Bennett u. a., 2008, S. 25 f.).

Deutlich wird einerseits, dass papierbasierte und computerbasierte Befragungen trotz gleicher Items nicht immer automatisch äquivalent sind. Andererseits zeigt die CTML, dass der Einsatz von Multimedia sehr zur Erklärung und zum Verständnis verschiedener Inhalte beitragen kann. Da durch die technische Entwicklung sich inzwischen die Formate in der computerbasierten Diagnostik

stark von denen im PP eingesetzten Itemformaten sowie früheren Formaten computerbasierter Diagnostik unterscheiden (können), ist entweder eine genaue Prüfung der Äquivalenz⁴² notwendig oder es ist von Beginn der Testkonstruktion zu entscheiden und berücksichtigen, in welchem Testmodus die Daten erhoben werden sollen. Des Weiteren sind Charakteristiken der potenziellen Zielgruppe einzubeziehen: Haben potenzielle Nutz und Nutzerinnen Erfahrung im Umgang mit Computern? Arbeiten sie gerne mit dem Computer? Stehen sie dem Einsatz von Computern in Testkontexten kritisch gegenüber? Der mögliche Einfluss dieser Faktoren (insbesondere der ICT-Literacy) auf die Performanz ist so weit wie möglich zu kontrollieren. Im Vorfeld ist daher bei jeglicher Gestaltung von computerbasierter Diagnostik auf eine zielgruppenadäquate Usability zu achten. Zusammenfassend fokussiert die Äquivalenzproblematik die Inhaltsvalidität von Datenerhebungsinstrumenten in unterschiedlichen Erhebungsformen. Im Rahmen dieses Forschungsvorhabens wird die Frage der Äquivalenz zwischen der papier- und der computerbasierten Diagnostik nicht weiter behandelt, da die Entwicklung und Evaluation hinsichtlich der Usability im Vordergrund stehen. Es wird zwar davon ausgegangen, dass durch eine ausgeprägte Usability auch die Inhaltsvalidität steigt. Doch ist die Überprüfung der Inhaltsvalidität erst nach Abschluss der Entwicklung möglich bzw. sinnvoll.

2.6 Usability

Bei der Gestaltung von Webseiten im Allgemeinen und computerbasierten Diagnoseinstrumenten im Besonderen sind neben zielgruppenspezifischen Voraussetzung und kognitionspsychologischen Aspekten. Die Anforderungen an eine ausgeprägte Usability werden über die Formulierung von Richtlinien konkretisiert. Bevor verschiedene Richtlinien für die Usability vorgestellt werden, wird der Begriff Usability mit seinen verschiedenen Bedeutungszuschreibungen vorgestellt, um schließlich eine Definition sowie Abgrenzung zu verwandten Begriffen vorzunehmen.

2.6.1 Usability – eine Begriffsan- und -einordnung

„Usability is a quality attribute that assesses how easy user interfaces are to use. The word "usability"

⁴² Guidelines der International Testing Commission (ITC) konstatieren, dass Äquivalenz zwischen computerbasierten und papierbasierten Tests dann besteht, wenn beide Tests vergleichbar reliabel sind, ausreichend hoch miteinander korrelieren, in ähnlichem Ausmaß mit anderen Testverfahren und externen Kriterien korrelieren sowie vergleichbare Mittelwerte und Standardabweichungen aufweisen (Jurecka & Hartig, 2007, S. 42).

also refers to methods for improving ease-of-use during the design process⁴³“ (Nielsen, 2012).

Der Begriff Usability wird im Deutschen meist mit Gebrauchstauglichkeit und/oder Benutzer-/Benutzerinnenfreundlichkeit übersetzt. Synonyme sind Benutzer-/Benutzerinnenfreundlichkeit, Bedienbarkeit und Brauchbarkeit. Allein die Existenz mehrerer Synonyme deutet auf eine nicht trennscharfe Verwendung des Begriffs im deutschen Sprachraum hin. Mit dem Begriff Gebrauchstauglichkeit wird stärker die Funktion und somit das Produkt, mit dem Begriff Benutzer-/Benutzerinnenfreundlichkeit stärker der Nutzer/die Nutzerin und somit eine Serviceorientierung fokussiert (Rampl, 2007). Da im Deutschen kein eindeutiges Äquivalent für den englischen Begriff Usability existiert, weil der englische Begriff Usability auch im Deutschen hinreichend geläufig ist und weil der Begriff Usability hinreichend definiert ist, wird im Folgenden der Begriff englischen Ursprungs *Usability* verwendet. Es existieren mehrere Definitionen des Begriffs Usability, die durch unterschiedliche Fokussierungen auch unterschiedliche Bedeutungsprägungen mitführen.

„Only by defining the abstract concept of „usability“ in terms of these more precise and measurable components can we arrive at an engineering discipline where usability is not just argued but is systematically approached, improved, and evaluated (possibly measured). [...] Clarifying the measurable aspects of usability is much better than aiming at a warm, fuzzy feeling of „user friendliness“ (Shackel, 1991, S. 24).

Um die Usability eines Produktes evaluieren und Usability-Anforderungen für eine bestimmte Zielgruppe entwickeln zu können, ist es unumgänglich, eine Definition herauszuarbeiten, die im Kontext dieser Arbeit Gültigkeit besitzt. Im Folgenden wird das Konstrukt Usability erörtert, um intersubjektiv nachvollziehbar dazustellen und zu begründen, welche Definition schließlich in der vorliegenden Arbeit Verwendung findet.

Der Begriff Usability wurde Mitte der 1980er Jahre eingeführt, um die Bezeichnung *user friendly* (engl. für nutzer-/nutzerinnenfreundlich) zu beschreiben. Der Anspruch einer möglichst erschöpfenden Bedeutung des Begriffs erfordert die Annäherung aus mehreren Richtungen. Wird der Wortstamm betrachtet,

⁴³ Das Zitat stammt von seiner Website <http://www.nngroup.com/articles/usability-101-introduction-to-usability/> (zuletzt geprüft am 03.11. 2014).

erhält er eine erste Schärfung: Der englische Begriff setzt sich aus den Begriffen *use* (engl. für *nutzen*) und *ability* (engl. für *Fähigkeit*, *Geschicklichkeit*, *Leistungsfähigkeit*) zusammen. Demnach bedeutet der Begriff, die *Fähigkeit etwas zu nutzen*. Diese Deutung impliziert, dass Webseiten oder Software ohne Usability sinnlos, weil nicht brauchbar, wären. Nach Rampl (2007) reicht allerdings das reine Erkennen der Notwendigkeit zu kompromisslosen Ausrichtung an den Bedürfnissen der Nutzer und Nutzerinnen nicht aus, sondern der Begriff impliziert ebenso eine Handlungsaufforderung: „*use the ability*“ (engl. für „Nutze die Möglichkeiten“). Durch den Imperativ ist somit nicht nur eine Serviceorientierung, wie der Begriff Benutzer-/Benutzerinnenfreundlichkeit vermuten lässt, sondern auch eine Notwendigkeit zu erkennen. Dies lenkt den Fokus etwas weg von der nutzer-/nutzerinnenzentrierten Perspektive und stärker hin zu einer funktionsorientierten Sichtweise.

Wird nun auch berücksichtigt, dass Normen und Standards für die Usability festgelegt sind, ist zudem eine weniger subjektiv abhängige und stärker objektiv feststellbare Ausprägung erkennbar. Somit prägen sowohl die Erfahrungen der Nutzer und Nutzerinnen als auch die Funktion der Technik die Ausprägung der Usability. An dieser Stelle lässt sich bereits erkennen, dass Usability ein relationaler Begriff ist: Die Usability ist immer vor dem Hintergrund des Gebrauchsziels der nutzenden Person zu betrachten: Ein Produkt dient dazu, dass ein Nutzer/eine Nutzerin eine bestimmte Aufgabe erfüllen möchte. Dabei kann es sich beispielsweise um eine Arbeitsaufgabe, ein Spiel oder einen Kaufprozess handeln. Schließlich laufen diese Erläuterungen auf vier Faktoren hinaus, die den Qualitätsgrad der Usability von Benutzer-/Benutzerinnenschnittstellen beeinflussen: a) Der Nutzer/Die Nutzerin, b) die Aufgabe, c) das System und d) die Umgebung bzw. der Anwendungskontext. So lautet eine frühe Definition von Usability:

„[Usability is] the ease of use and acceptability of a system or product for a particular class of users carrying out specific tasks in a specific environment; where ‘ease of use’ affects user performance and satisfaction, and ‘acceptability’ affects whether or not the product is used“ (Bevan, Kirakowski & Maissel, 1991, S. 2).

Usability ist die Einfachheit der Nutzung und die Akzeptanz von einem Produkt, welche Auswirkungen auf die Effizienz und Zufriedenheit einer Produktnutzung haben; die Akzeptanz hat wiederum Einfluss darauf, ob das Produkt genutzt wird. Diese Definition macht insbesondere die Relationalität des Konstrukts Usability deutlich.

Für eine weitere Priorisierung der Faktoren innerhalb des Konstrukts kann die Definition von Nielsen herangezogen werden. Nielsen (1993) geht von der *Usefulness* (engl. für *Brauchbarkeit*) aus, welche beschreibt, ob ein System genutzt werden kann, um ein bestimmtes Ziel zu erreichen. Die *Usefulness* lässt sich wiederum in *Utility* und *Usability* unterteilen. *Utility* bezeichnet dabei, ob ein System für die Anliegen genutzt werden kann, für die es gedacht ist. Die *Usability* ist wiederum der Grad an Qualität, in welcher der Nutzer/die Nutzerin die Interaktion mit etwas erlebt (Nielsen, 1993, S. 25). In diesem Kontext sind die Attribute für die Qualität Erlernbarkeit, Effizienz, Erinnerbarkeit, Fehlertoleranz und Zufriedenheit (vgl. zur näheren Beschreibung der Attribute Kap. 2.6.4). Diese Bezeichnung bezieht – im Gegensatz zu der deutschen Übersetzung – das Erleben des Nutzers/der Nutzerin stärker ein. *Usability* ist somit keine eindimensionale Eigenschaft eines Nutzer-/Nutzerinneninterfaces.

Die *International Organization for Standardization* (ISO) und das Deutsche Institut für Normung (DIN) geben für *Usability* folgende Definition vor⁴⁴:

„[*Usability* ist] das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und mit Zufriedenheit zu erreichen“ (DIN EN ISO, 2011, S. 38).

Diese Definition geht über die von Nielsen (s.o.) hinaus. Nielsen legt den Fokus auf die Leichtigkeit und Unkompliziertheit der Produktnutzung. Der DIN Norm ist diese Anforderung implizit, sie liefert darüber hinaus ein differenziertes Bild, indem Kontextabhängigkeit (Nutzungskontext), die Nutzbarkeit der Technologie (effektiv und effizient) und die Emotionalität des Nutzers/der Nutzerin berücksichtigt werden. Die *Usability* bezieht sich auf die Zeitspanne während der Nutzung eines Produktes. Das bedeutet nicht, dass die *Usability* während der Entwicklung – also vor der Nutzung – keine Rolle spielt, sondern dass die *Usability* nur während der Nutzung beobachtbar und messbar ist.

Da die Definition der ISO bzw. DIN einerseits hinreichend differenziert und andererseits auch eine der geläufigsten ist, wird dem Begriff *Usability* die Definition der ISO bzw. des DIN im weiteren Verlauf der Arbeit zu Grunde gelegt. Ein weiterer Begriff, der fälschlicherweise in manchen Fällen mit *Usability* gleichgesetzt wird, ist der Begriff *Software-Ergonomie*. Die *Software-Ergonomie* ist ein der *Usability* übergeordneter Begriff und schließt diese ein. Die *Software-Ergonomie* besteht aus den drei Bereichen *Gestaltungsgrundsätze*

⁴⁴ Die Definition der ISO und des DIN werden in Kap. 2.6.4 wieder aufgegriffen und eingehend erläutert.

und Rahmenbedingungen, Multimedia-Navigation und Steuerung sowie Auswahl und Kombination relevanter Medien (die Anforderungen an die Software-Ergonomie wurden vom Institut für Normung in der Richtlinie ISO 14915 formuliert). Hier gerät wiederum die Perspektive des Nutzers/der Nutzerin an den Rand des Blickfeldes.

Die Erläuterung des Begriffs Usability sowie die für die Usability relevanten verwandten Konzepte macht deutlich, dass der Begriff Usability einerseits nicht isoliert betrachtet werden kann. Andererseits ergeben sich sowohl Überschneidungen als auch hierarchische Beziehungen zwischen den Konzepten. Die Abbildung 3 dient dazu, das Beziehungsgeflecht der Begrifflichkeiten zu verdeutlichen. Die Abbildung 3 kann nicht den Anspruch der Vollständigkeit erfüllen, wohl aber als Gerüst für den weiteren Verlauf dieser Arbeit dienen.

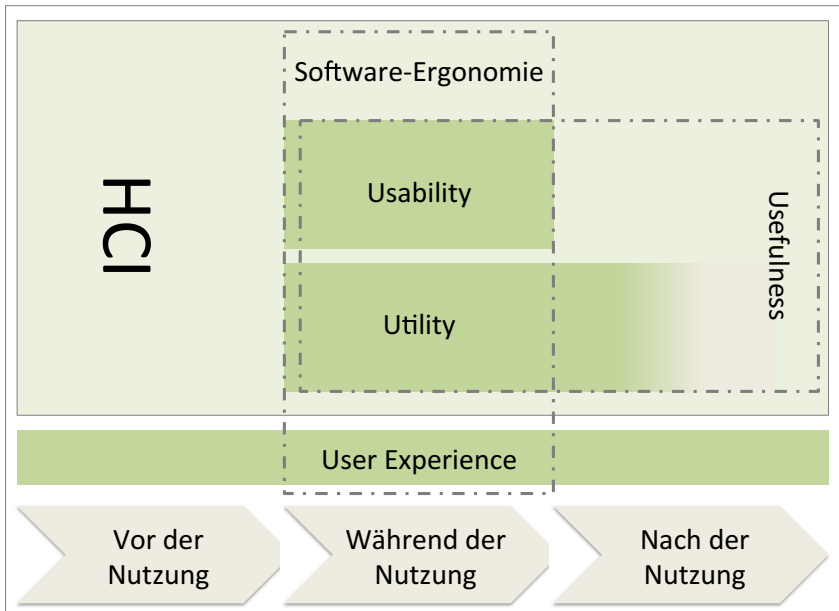


Abbildung 3: Usability in Beziehung zu angrenzenden/verwandten Begriffen

In dieser Abbildung sind die Begriffe unter Berücksichtigung des Zeithorizonts in Beziehung gesetzt. Die Usability ist ein Konstrukt der Human-Computer Interaction (HCI) (Mensch-Computer-Interaktion), in der es um benutzer-/benutzerinnenfreundliches Design von Computersoftware oder Webseiten geht (Preece, 1994). Die HCI bezieht sich auf den Zeitraum vor, während und nach

der Nutzung eines Produktes. Ein Bestandteil der HCI ist die Software-Ergonomie, welche die Begriffe *Usability* und *Utility* einschließt. *Usability* und *Utility* ergeben zusammen die *Usefulness* eines Produktes. Die *Utility* und *Usefulness* gehen über den gegenwärtigen Zeitpunkt der Nutzung hinaus, indem auch der längerfristige – zukünftige Nutzen – mitgedacht wird.

Ein Aspekt, der ebenfalls über den gegenwärtigen Zeitpunkt hinaus geht und zusätzliche auch die Vergangenheit einschließt, ist die *User Experience* (UE). Diese gibt Aufschluss über die Wahrnehmung eines Produkts aus der Perspektive der Nutzer und Nutzerinnen und wird oftmals in Zusammenhang mit der *Usability* überprüft. Mit Hilfe von Erkenntnissen über die UE kann die *Usability* eines Produktes gesteigert werden. Kenntnisse über die UE sind insbesondere hilfreich, wenn über die potenzielle Nutzer-/Nutzerinnengruppe hinsichtlich der Computernutzung und -erfahrung wenig bekannt ist. Da dies bezüglich der Forschungsfrage der Fall ist, soll die UE einer genaueren Betrachtung unterzogen und mit der *Usability* in Beziehung gesetzt werden.

2.6.2 *User Experience (UE)*

Ins Deutsche übersetzt bedeutet UE zunächst Nutzer-/Nutzerinnenerfahrung. Sie wird definiert als „Wahrnehmungen und Reaktionen einer Person, die aus der tatsächlichen und/oder der erwarteten Benutzung eines Produktes, eines Systems oder einer Dienstleistung resultieren" (DIN EN ISO, 2009) und ist auf *interaktive* Computersysteme bezogen (Herczeg, 2009, S. 7).

Damit liegt der Fokus auf den Erfahrungen der Nutzer und Nutzerinnen hinsichtlich eines Produkts. Die Erfahrungen sind allerdings nicht nur abhängig von dem gegenwärtig genutzten System und der zu bearbeitenden Aufgabe, sondern auch von den Erfahrungen, welche der Nutzer/die Nutzerin mitbringt (z. B. haben Nutzer und Nutzerinnen, die viele positive Erfahrungen mit Computern gemacht haben sicherlich ein anderes Erleben während der Nutzung eines Produktes als Personen, die bisher wenig und negative Erfahrungen gemacht haben). Sie ist insbesondere zu berücksichtigen, wenn ausschließlich der Nutzer/die Nutzerin über den Verbleib auf oder das Verlassen einer Website entscheidet, also freiwillig ein Produkt nutzt (Rampl, 2007).

Die UE bekam durch den vermehrten Einsatz in unterschiedlichsten Lebenssituationen eine zunehmende Bedeutung. Die Nutzer und Nutzerinnen erwarteten mehr als ein funktionierendes System, nämlich Unterhaltung. Die UE ist einer der neuesten Begriffe in dem Kriteriensatz, nach welchem ein System evaluiert werden soll (Petrie & Bevan, 2009, S. 5). Erstmals findet sich der Begriff UE 1998 in einer Veröffentlichung von Norman, der die Berücksichtigung von Aspekten verlangte, die über Effektivität, Effizienz und Zufriedenheit hinausgehen (Norman, 1998).

Die UE umfasst demnach alle Effekte auf den Nutzer/die Nutzerin vor und nach der Nutzung des Produktes. Die Usability hingegen bezieht sich nur auf die Zeitspanne während der Nutzung. Die vorherigen Erfahrungen und die Emotionalität des Nutzers/der Nutzerin werden – wie bereits oben erläutert – weniger berücksichtigt. So kann beispielsweise die UE hoch, die Usability aber niedrig sein. Aufgrund der Einflussfaktoren Mensch, Technologie, Aufgabe und Anwendungskontext ist auch die UE – wie Usability – ein relationaler Begriff.

Schlussfolgernd hängt von den Fähigkeiten und der Wahrnehmung der Nutzer und Nutzerinnen, der Technologie, der zu bearbeitenden Aufgabe und dem Anwendungskontext ab, welche Erfahrungen die Nutzer und Nutzerinnen mit einem Produkt machen und ob ein Produkt als gebrauchstauglich und nutzer-/nutzerinnenfreundlich bewertet wird. Da in dieser Arbeit der Fokus auf den Nutzern und Nutzerinnen liegt und der Frage nachgegangen wird, welche zielgruppenspezifischen Anforderungen hinsichtlich der Usability ein Produkt erfüllen sollte, wird im Folgenden näher erläutert, welche menschlichen Faktoren bei der gebrauchstauglichen und benutzer-/benutzerinnenfreundlichen Gestaltung zu berücksichtigen sind.

2.6.3 Kognitive und physiologische Voraussetzungen

Die kognitiven und physiologischen Voraussetzungen bestimmen, wie Menschen Informationen individuell wahrnehmen und verarbeiten. Die Informationen werden vorerst physisch wahrgenommen und ausgehend von den individuellen mentalen Modellen kognitiv verarbeitet. Die kognitiven und physiologischen Voraussetzungen stellen somit eine zu beachtende Einflussgröße im Kontext der Usability dar.

Im Kap. 2.4 wurde mit der CLT bereits erläutert, dass die Auslastung des Arbeitsgedächtnisses sowohl von der Darbietungsform als auch vom Inhalt abhängig ist. Ob Informationen das Arbeitsgedächtnis stark oder weniger stark auslasten hängt davon ab, wie anstrengend und schwierig es für eine Person ist, sich den Inhalt zu erschließen. Grundlegend wird dies von den physiologischen und kognitiven Voraussetzungen bestimmt. Zu den kognitiven Prozessen zählen u. a. die Wahrnehmung, die Aufmerksamkeit, das Gedächtnis sowie Wissensrepräsentationen und mentale Modelle⁴⁵. Bei der Nutzung von Software bestehen nach Norman (1996) zwei wesentliche Hürden: Der *gulf of execution* (Ausführungskluft) und der *gulf of evaluation* (Auswertungskluft) (Norman, 1996). Der *gulf of execution* beschreibt die Situation, dass der Nutzer/die Nutzerin dem System seine Intention vermitteln muss – beispielsweise mit Hilfe der Eingabe

⁴⁵ Weitere kognitive Prozesse – beispielsweise höhere Hirnleistungen wie Rechnen und Erinnern – werden nicht weiter thematisiert, da diese nicht zwangsläufig eine Rolle bezüglich der Usability spielen.

über die Tastatur oder das Anklicken mit der Computermaus. Bei dem *gulf of evaluation* geht es darum, dass die Ausgaben des Systems von der nutzenden Person korrekt interpretieren und in ihre Handlungen integrieren muss. Für Beides sind sowohl kognitive als auch physiologische Voraussetzungen zu berücksichtigen.

Im Folgenden werden die kognitiven und physiologischen Voraussetzungen sowie deren Rolle bezüglich der Usability kurz erläutert. Bei der Darstellung der kognitiven und physiologischen Grundlagen wird von den Voraussetzungen bei Personen ausgegangen, die weder kognitiv noch physisch eingeschränkt sind.

2.6.3.1 Aufmerksamkeit

Das Thema Aufmerksamkeit wird kontrovers diskutiert (Sarodnick & Brau, 2011, S. 57) und soll hier nur so weit erläutert werden, wie es zum weiteren Verständnis dient.

Aufmerksamkeit ist die auf die Betrachtung eines Objekts gerichtete Bewusstseinshaltung, durch die das Beobachtungsobjekt bewusst wahrgenommen wird (Bergius, 2004, S. 84f). Die Aufmerksamkeit ist begrenzt und legt fest, welche Ereignisse uns bewusst werden. Beispielsweise kann zwischen visueller, auditiver und zentraler Aufmerksamkeit unterschieden werden (J. R. Anderson, 2013), was zu Folge hat, dass sie situationsabhängig auf einzelne Reize (selektive Aufmerksamkeit) oder auf mehrere Reize gleichzeitig gelenkt (geteilte Aufmerksamkeit) wird.

Bezüglich der visuellen Aufmerksamkeit ist die Beschränkung einfacher ersichtlich, denn wir können nur auf eine Stelle zur Zeit blicken und diese deutlich wahrnehmen. Allerdings ist damit nicht unbedingt festgelegt, dass unsere Aufmerksamkeit ausschließlich auf diesen Punkt gerichtet ist. Die Spotlight-Metapher dient einer bildlichen Beschreibung: Die Aufmerksamkeit wird wie ein Scheinwerfer auf verschiedene Bereiche gelenkt und je nachdem, wie der Scheinwerfer eingestellt ist, wird ein Punkt sehr hell erleuchtet (scharf gesehen) oder ein größerer Bereich in Augenschein genommen (zu Lasten der Sehschärfe) (vgl. J. R. Anderson, 2007, S. 100). Ungeübte Prozesse benötigen stärkere Aufmerksamkeit als automatisierte Prozesse. Beispielsweise ist das Lesen eines Wortes für funktionale Alphabeten und Alphabetinnen nicht automatisiert und erfordert eine größere Aufmerksamkeit als für Personen, die sich auf einem höheren Kompetenzniveau bezüglich des Lesens befinden.

Erfolgen zwei automatisierte Prozesse parallel, können trotz der Automatisierung Verzögerungen auftreten, weil zwei Informationen nicht gleichzeitig verarbeitet werden können. So zeigt z. B. der Stroop-Effekt, dass es Menschen schwer fällt, die Druckfarbe eines Wortes (z. B. blau) zu benennen, wenn es sich bei dem Wort um eine andere Farbe handelt (z. B. rot). Das Lesen eines Wortes

ist ein automatisierter Prozess, so dass es schwierig ist, ihn zu unterdrücken. Der automatisierte Prozess interferiert mit der Verarbeitung der anderen Informationen, die sich auf das Wort beziehen. Interferenzen können zudem auftreten, wenn zwei Reaktionen aufeinander folgen. Der Konflikt ist umso größer, je stärker die beiden Reaktionen auf die gleichen Verarbeitungsressourcen zurückgreifen (J. R. Anderson, 2007, S. 124).

Für die Gestaltung eines Systems lässt daraus ableiten, dass die Aufmerksamkeit durch akustische Signale oder visuelle Hervorhebungen gesteuert werden kann. Das kann zudem einen unterstützenden Effekt auf die Wahrnehmung haben.

2.6.3.2 Wahrnehmung

Die Wahrnehmung bezeichnet den Prozess und das Ergebnis der Informationsverarbeitung und steht dabei im Zusammenhang mit der Aufmerksamkeit: Die wahrnehmende Person muss entscheiden, worauf sie Aufmerksamkeit richtet. Dafür müssen die sensorischen Systeme erkennen, um was es sich in der Außenwelt handelt. Übertragen auf Computer und Interfaces ist die Frage, wie das Gehirn Symbole und Darstellungen auf dem Computerbildschirm erkennt und identifiziert. Bezüglich der Wahrnehmung ist grundlegend zwischen dem fovealen und dem peripheren Sehen zu unterscheiden (J. R. Anderson, 2007, S. 51). Bei Ersterem handelt es sich um die Reizaufnahme über einen kleinen zentralen Bereich der Netzhaut in dem scharf gesehen wird. Das foveale Sehen ist beispielsweise Voraussetzung für das Lesen. Beim peripheren Sehen werden umliegende Reize entdeckt, welche die Augenbewegung steuern.

Für die Gestaltung von Benutzer-/Benutzerinnenoberflächen spielen insbesondere die 1. *Objektwahrnehmung* und die 2. *visuelle Mustererkennung* eine Rolle:

1. *Objektwahrnehmung*: Visuelle Reize und Szenen werden mit Hilfe der sog. Gestaltprinzipien in Objekte gegliedert (J. R. Anderson, 2007, S. 60). Nach Wertheimer (1912) werden die visuellen Reize nach den vier Gestaltesetzen zu Objekten zusammengeführt.

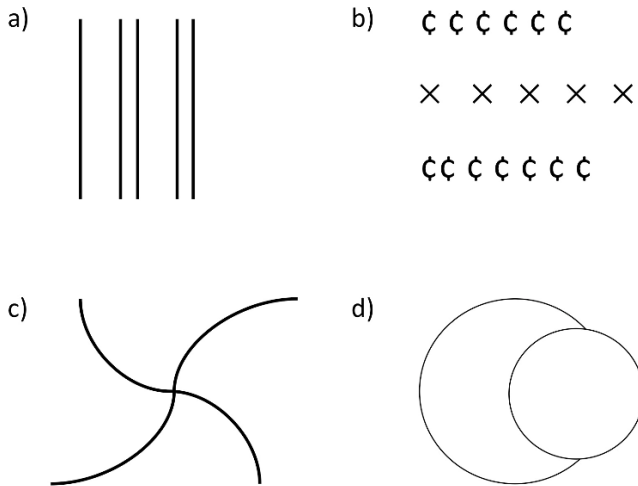


Abbildung 4: Grafische Darstellung der vier Gestaltgesetze in Anlehnung an Anderson (2007, S. 60)

- Das *Gesetz der Nähe* besagt, dass naheliegende Objekte als zusammengehörig empfunden werden.
- Tritt das *Gesetz der Ähnlichkeit* in Kraft werden ähnlich aussehende Objekte als Gruppe wahrgenommen.
- Beim *Gesetz des glatten Verlaufs* werden Linien so zusammengehörig wahrgenommen, dass sich ein glatter Verlauf anstatt von Ecken ergibt.
- Das *Gesetz der Geschlossenheit* bzw. der guten Gestalt drückt aus, dass Objekte als überdeckend wahrgenommen werden, wenn sich daraus eine geschlossene, klare Form ergibt, was z. B. eine räumliche Wahrnehmung hervorrufen kann.

Wertheimer beobachtete in Experimenten, in denen Striche auf verschiedene Weise angeordnet waren, dass Personen zwar nicht glauben, sie bewegen sich tatsächlich aufeinander zu doch sie sehen eine Tendenz der Striche zueinander „hin“. Je stärker die Aufmerksamkeit auf die beiden Objekte gerichtet war, desto stärker trat das Phänomen auf (Wertheimer, 1912, S. 226).

Die Gestaltgesetze können bei der Erstellung einer grafischen Oberfläche berücksichtigt werden, z. B. die Gesetze der Nähe und der Ähnlichkeit bei der Gestaltung eines Navigationsmenüs: Einerseits kann durch die aneinander liegende Anordnung der Menüpunkte bzw. -buttons zum Ausdruck gebracht werden, dass diese zu einer Kategorie gehören. Andererseits wird auch durch die

ähnlich Gestaltung der Schaltflächen der Eindruck erzeugt, dass die Objekte zu einer Gruppe (z. B. dem Webseitenmenü) gehören.

Neben der Gestaltung und Anordnung von Objekten kann auch der zeitliche Aspekt einen Einfluss auf die Wahrnehmung haben. Wenn Schaltflächen blinkend dargestellt werden, kann die Frequenz des Aufleuchtens tempoabhängig die Aufmerksamkeit auf sich ziehen. Dies trifft ebenfalls zu, wenn Zielreize durch Farbe, Größe oder Animation hervorgehoben werden. Wie bereits in Kap. 2.4 erläutert wurde, ist der *Extraneous Load* je nach Darstellung und Menge der Informationen ausgelastet. So kann auch für die Wahrnehmung gefolgert werden, dass für die Verarbeitung von Reizen je nach Menge und Darstellung die Gedächtniskapazität für die Wahrnehmung und Verarbeitung unterschiedlich stark beansprucht wird.

2. Visuelle Mustererkennung: Visuelle Informationen können in Objekte zergliedert werden, doch um die Welt sehen zu können, müssen die Objekte auch identifiziert werden. Dies geschieht mit der Mustererkennung. Sie liefert die Erklärung dafür, wie zum Beispiel die Darbietung des Buchstabens A auch als eine Ausprägung des Musters A erkannt wird. Eine Möglichkeit zur Erkennung ist der sogenannte Schablonenabgleich, bei dem ein Netzhautbild des Objektes an das Gehirn übermittelt wird, um es mit bereits gespeicherten Mustern zu vergleichen (J. R. Anderson, 2007, S. 61). Nach dieser Annahme versucht das Gehirn, das Bild eines Buchstabens mit jeder Schablone im Gehirn, die für Buchstaben vorhanden sind, abzugleichen und die Schablone mit der besten Übereinstimmung zu melden. Eine weitere Betrachtungsweise geht davon aus, dass die Mustererkennung auf einer Merkmalsanalyse beruht. Demnach wird jeder Reiz als Kombination elementarer Merkmale angesehen. Übertragen auf das Beispiel der Buchstabenerkennung sind dies die horizontalen, vertikalen, schrägen und/oder gekrümmten Linien der Buchstaben.

Bei der Wahrnehmung von Objekten werden einfache Teilobjekte (wie z. B. der lange Hals, vier Beine und ein Schwanz bei einer Giraffe) zu einem Gesamtobjekt zusammengeführt. Vertraute Objekte werden als Konfigurationen einzelner einfacher Komponenten zusammengefasst.

Schlussfolgernd sind Informationen entsprechend der Erkenntnisse zur Wahrnehmung so darzustellen, dass sie schnell erkannt bzw. erschlossen werden können. Starke Abweichungen von gängigen und geläufigen Mustern kann zu Irritationen, verlängerten Verarbeitungszeiten sowie einer starken Beanspruchung des *Extraneous Load* führen.

2.6.3.3 Wissensrepräsentation und Mentale Modelle

Nachdem Objekte wahrgenommen und erkannt (assimiliert) bzw. neue Objekte erschlossen (akkomodiert) wurden, werden die Informationen im Gehirn weiter

verarbeitet. Die Weiterverarbeitung von Informationen hängt davon ab, auf welche Weise sie im Gehirn repräsentiert sind (J. R. Anderson, 2007, S. 129). Grundlegend kann zwischen wahrnehmungsbasierter und bedeutungsbezogener Wissensrepräsentation unterschieden werden (J. R. Anderson, 2013, S. 97). Erstere bezieht sich auf direkt wahrgenommene Reize und ihre Repräsentationen im Gehirn. Bei der bedeutungsbezogenen Wissenspräsentation werden die Informationen abstrahiert und bezüglich ihrer Bedeutung betrachtet. Der Nutzer/Die Nutzerin beurteilt auf der Basis des mentalen Modells die Folgen seiner/ihrer beabsichtigten Handlung. Mentale Modelle können durch Lernen und Erfahrungen permanent (weiter)entwickelt werden⁴⁶ (Niegemann, 2008, S. 284).

Im Kontext der HCI entwickeln Nutzer und Nutzerinnen über Eigenschaften und Funktionsweisen der Mensch-Computer-Interaktion und der Technologie mentale Modelle (Niegemann, 2008, S. 284). Auch bei der Verarbeitung visueller Informationen werden mentale Bilder verwendet (J. R. Anderson, 2007, S. 134). Sie werden bei Bedarf aus dem Gedächtnis aufgerufen. Wenn eine Person beispielsweise ein Computerprogramm aufruft, erzeugt die Person vor der eigentlichen Handlung (dem Anklicken des Programms) und dem eigentlichen Erscheinen des Programms – basierend auf ihrem bisherigen Wissen sowie bisherigen Erfahrungen mit Computerprogrammen und dessen Funktionsweise – eine mentale Vorstellung der Eigenschaften und Interaktionsmöglichkeiten. Mentale Modelle basieren also auf der subjektiven Wahrnehmung und nicht auf Fakten. Sie sind somit individuell – zwei Nutzer/Nutzerinnen können somit auch unterschiedliche mentale Modelle von einem Interface haben. Die Eignung eines mentalen Modells für die Arbeit mit einem Computerprogramm ist davon abhängig, wie systematisch und korrekt ein solches Modell entstanden ist (Herczeg, 2009, S. 52). Die Konstruktion mentaler Modelle von räumlichen Sachverhalten gelingt allerdings anhand von Texten weniger gut als anhand von Bildern (Brünken, Steinbacher, Schnotz & Leutner, 2001). Beispielsweise ist eines der größten Dilemmas, dass eine Lücke zwischen den gängigen mentalen Modellen von Interfacedesignern und den Nutzern/Nutzerinnen besteht. Begründet ist dies meist darin, dass Designer über zu viele Informationen verfügen und daher komplexere mentale Modelle entwickeln, die zu einer komplexeren Navigationsstruktur führen (Nielsen, 2010).

Da bildhaftes Material besser behalten werden kann als verbales (Brünken u. a., 2001; Sarodnick & Brau, 2011, S. 59), sind Symbole und Icons vermutlich besser geeignet, um einerseits mentale Modelle zu bilden als auch um mentale Modelle abzurufen. Um Informationen zu verarbeiten ist allerdings die duale

⁴⁶ Die Weiterentwicklung von mentalen Modellen wird als Assimilation, die Neuentwicklung von mentalen Modellen als Akkomodation bezeichnet (Graf, Zimbardo & Gerrig, 2007, S. 66). Diese Bezeichnungen gehen auf Jean Piaget zurück (vgl. z. B. Piaget, Aebli & Montada, 2003).

Repräsentation – visuell und verbal – effektiver. So erbringen Menschen bessere Gedächtnisleistungen, wenn Informationen visuell als auch verbal enkodiert werden (J. R. Anderson, 2007; Brünken u. a., 2001; Paivio, 1971).

Eine besondere Relevanz haben im Kontext der Nutzung von Programmen durch Personen mit niedriger Literalität Symbole und Icons, da diese ggf. die einzige Möglichkeit darstellen, mentale Modelle abzurufen. Wörter wären dafür weniger geeignet; erstens greifen nur mit einer geringen Wahrscheinlichkeit (in Abhängigkeit der Wortschwierigkeit) Automatismen (vgl. Kap. 2.6.3.3) und zweitens nimmt das Erfassen eines Wortes Zeit in Anspruch, erfordert hohe kognitive Ressourcen oder ist gar nicht erst möglich, so dass kein Sinn und somit auch kein mentales Modell abgerufen werden kann. Schlussfolgernd bieten Symbole das Potenzial, leichter mentale Modelle abzurufen aber auch falsch zu interpretieren. Werden beispielsweise erst mentale Modelle nach Aufrufen eines Symbols im Kontext eines Programmes gebildet, ist das mentale Modell mit dem programmspezifischen Symbol verknüpft und in seiner Funktionsweise richtig interpretiert. Werden Symbole aber uneinheitlich verwendet, kann das zu Konfusionen führen.

Kurz zusammengefasst ist die Wahrnehmung die Grundlage für die Aufnahme und Verarbeitung von Informationen. Symbole, Texte und Sprache müssen daher verständlich aufbereitet werden. Zudem sind Symbole so vereinfacht darzustellen, dass sie leicht im Gedächtnis bleiben und wiedererkannt werden können. Umgekehrt sind aber auch geeignete Eingabemöglichkeiten anzubieten. Für die Interaktion mit dem Medium Computer, z. B. über die Tastatur und Computermaus, sind daher auch physiologische Voraussetzungen relevant.

2.6.3.4 Physiologische Voraussetzungen

Zu den physiologischen Voraussetzungen bezüglich des Umgangs mit Computern zählen motorische sowie sensorische Fähigkeiten. Unter Motorik ist die mechanische Ausführungen von Handlungen zu verstehen. Motorische Fähigkeiten beziehen sich insbesondere auf den Tastsinn und die kinästhetische Sensitivität. Hierzu zählt auch die Bewegungskontrolle, welche die Handlung aufgrund von kognitiver und perceptiver Steuerung bezeichnet (Fisk, Rogers, Charness, Sharit & Czaja, 2009, S. 15). Bezüglich der HCI ist die Motorik die wesentliche Methode, um auf das System einzuwirken. Die motorischen Fähigkeiten können beispielsweise einen Einfluss auf die präzise Steuerung der Hände haben und somit den Umgang mit der Maus, die Platzierung des Mauszeigers auf den Schaltflächen und damit die Aktivierung von Schaltflächen (das Anklicken) erschweren. Auch kann durch motorische Einschränkung die Eingabe über die Tastatur beeinträchtigt sein. Im Laufe des Alterns nehmen diese Fähigkeiten

sukzessive ab⁴⁷. Beispielsweise reagieren ältere Menschen tendenziell langsamer (ältere Menschen reagieren durchschnittlich 1,5 bis zweimal langsamer als jüngere Erwachsene) und Bewegungen werden zunehmend unpräziser. Auch die Sensualität – das Empfindungsvermögen insbesondere bezüglich der Sinne Sehen, Hören, Schmecken, Riechen und Fühlen – nimmt tendenziell ab dem 60. Lebensjahr ab (Fisk u. a., 2009, S. 26). Allerdings spielen Riechen und Schmecken bei dem momentanen Stand der Technik im Usability-Kontext keine Rolle (wobei nicht auszuschließen ist, dass zukünftig auch diese Sinne via Computer angeregt werden können). Die Sinne Sehen und/oder Hören werden am Computer hingegen permanent angesprochen. Die Informationen mit Computern werden primär über den visuellen Kanal aufgenommen (z. B. bei Webseiten ohne auditive Unterstützung). Häufig werden aber auch gleichzeitig der visuelle und der auditive Kanal angesprochen (z. B. durch Filme). Die Wahrnehmung der Informationen hängt somit von der visuellen und auditiven Fähigkeit des Menschen ab, wobei insbesondere die Seh- und Hörfähigkeit mit zunehmendem Alter abnehmen (Sharit, Fisk, Rogers, Charness & Czaja, 2007, S. 17). Für die Usability und die UE sind diese heterogenen Voraussetzungen hinsichtlich der angenommenen Zielgruppe zu berücksichtigen. Ist beispielsweise davon auszugehen, dass in der Zielgruppe vermehrt Personen mit eingeschränkten motorischen Fähigkeiten vorhanden sind, sollten interaktive Schaltflächen größer im Verhältnis zur „Norm“ gestaltet werden. Zudem können Funktionen eingebunden werden, welche die Anpassung an individuelle Bedürfnisse ermöglichen, wie z. B. die Einstellungsmöglichkeiten von Lautstärke und Schriftgröße.

Werden zusammenfassend wieder die Darlegungen zu kognitiven und physiologischen Voraussetzungen mit den drei Einflussfaktoren Mensch, Technologie und Aufgabe in Beziehung gesetzt, so haben die menschlichen kognitiven und physiologischen Voraussetzungen einen Einfluss darauf, wie das Interface vom Menschen wahrgenommen wird, wie die Informationen verarbeitet werden und wie mit dem System interagiert wird. Diese Einflussfaktoren sind wiederum bei der Gestaltung der Technologie und der Aufgabe zu berücksichtigen und bilden die Grundlage für die Anforderungen an die Usability einer Software.

2.6.4 Allgemeine Usability-Anforderungen und Heuristiken

Es existieren unzählige Handbücher zum Thema Usability. Zudem existieren verschiedene Richtlinien (engl. *Guidelines*), Prinzipien, Grundsätze und Heuristiken, die dazu dienen sollen, einen hohen Grad an Usability sicherzustellen.

⁴⁷ In Deutschland ist über alle Alphastufen hinweg der Anteil funktionaler Analphabeten und Analphabetinnen bei den 50 bis 64jährigen Personen am größten (Grotlüschen & Riekmann, 2011, S. 7).

Neben dem internationalen Standard (EN ISO 9241) existieren weitere Qualitätskriterien und Heuristiken, die für den Großteil der Nutzer und Nutzerinnen Gültigkeit besitzen (vgl. z. B. Nielsen, 2012; Sarodnick & Brau, 2011; Shneiderman & Plaisant, 2005). Zudem bestehen Richtlinien für besondere Zielgruppen, z. B. ältere Personen (Fisk u. a., 2009) und Personen mit körperlichen und/oder geistigen Einschränkungen (Caldwell u. a., 2008). Der Unterschied zwischen Richtlinien und Heuristiken besteht darin, dass Richtlinien (im deutschen Sprachraum wird auch häufig das englische Wort *Guidelines* verwendet) stärker konkretisiert sind als Heuristiken.

Der Web Content Accessibility Guidelines (WCAG) (Richtlinien für barrierefreie Webinhalte) bestehen beispielsweise aus ca. 60 Richtlinien. Das United States Dept of Health and Human Services extrahierte auf der Grundlage von über 500 Richtlinien schließlich 209, die nach relativer Wichtigkeit und Aussagekraft beurteilt wurden. Diese sind in 17 Bereiche kategorisiert (z. B. UE, Texterscheingung, Accessibility und Seitenlayout) und bestehen wiederum aus mehreren konkret formulierten Richtlinien.

Die Richtlinien des DIN und des United States Dept of Health and Human Services sind in Abschnitte eingeteilt, die auch als grundlegende Prinzipien oder Heuristiken bezeichnet werden. Heuristiken sind leichter handhabbar, doch sind sie durch ihren Abstraktionsgrad von dem jeweiligen Nutzungskontext und der Interpretation der Experten und Expertinnen sowie deren Kompetenz abhängig (Hertzum, 2010, S. 571). Eine Heuristik ist die Lehre oder Wissenschaft von den Verfahren, um Probleme zu lösen mit Hilfe von methodischen Anleitungen und Anweisungen zur Gewinnung neuer Erkenntnisse. Sie besteht aus vereinfachten Annahmen, mit deren Hilfe ein Problem schneller gelöst werden kann als ohne Vereinfachung. Die Nachteile einer solchen Vereinfachung bestehen darin, dass eine Heuristik zwar schnell realisierbar ist, aber nicht mit Sicherheit zur korrekten Lösung führt (Gegensatz zu Lösungsalgorithmus) (Zimmer, 2004, S. 400).

Für die Überprüfung und Evaluation der Onlinetestumgebung für funktionale Analphabeten und Analphabetinnen werden Usability-Heuristiken als Evaluationsgrundlage herangezogen. Da bisher keine Erkenntnisse im Bereich der Usability-Anforderungen existieren und unter Berücksichtigung der vorhandenen Ressourcen für die Entwicklung und Evaluation ist eine Annäherung an eine erfüllte Usability über „grobe“ Heuristiken effizienter als über detaillierte Richtlinien. Im Folgenden werden verschiedene Usability-Heuristiken vorgestellt, kritisch reflektiert und zusammengefasst.

2.6.4.1 Usability-Heuristiken

Vorgestellt werden im Folgenden solche Heuristiken, die am stärksten etabliert (EN ISO 9241), viel zitiert (Nielsen, 2012), stark konkretisiert (Leavitt & Shneiderman, 2006) sind und besondere zielgruppenspezifische Charakteristika berücksichtigen, die auch auf funktionale Analphabeten und Analphabetinnen übertragbar sind (z. B. Fisk u. a., 2009). Die Heuristiken der DIN-Norm werden dabei am stärksten fokussiert, da sie sich am stärksten etabliert haben und im aktuellen Diskurs zur Usability-Forschung einen wesentlichen Referenz- und Bezugspunkt darstellen.

Die EN ISO 9241 ist der internationale Standard, welcher die Anforderungen an die Ergonomie der Mensch-System-Interaktionen beschreibt. Vorerst betitelte die International Standard Organisation (ISO) die Richtlinie mit „Ergonomische Anforderungen für die Bürotätigkeit mit Bildschirmgeräten“. Im Jahr 2006 wurde der Titel jedoch in „Ergonomie der Mensch-System-Interaktion“ umbenannt, um nicht nur auf Bürotätigkeit bezogen zu sein.

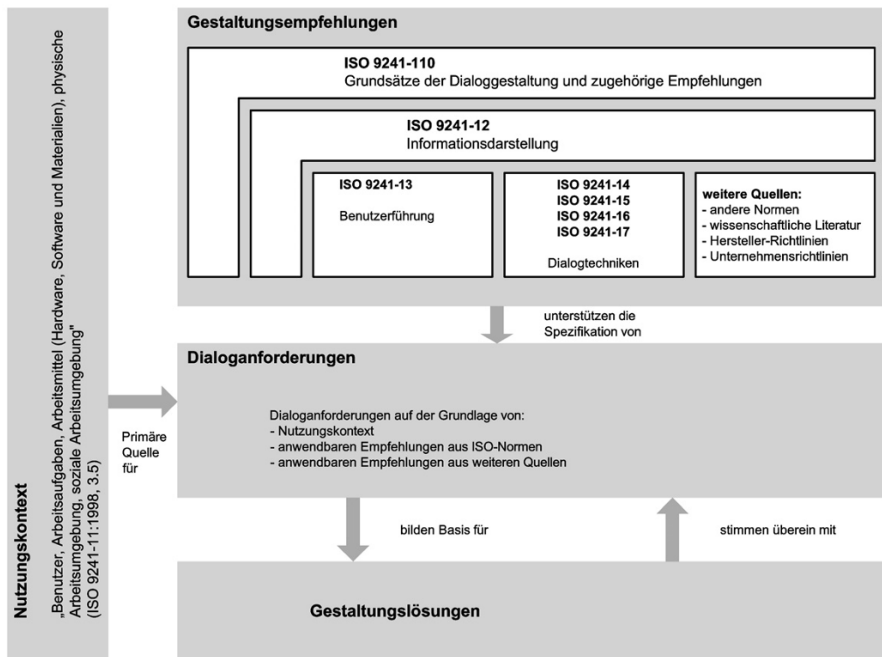


Abbildung 5: Gestaltungs-/Anwendungsrahmen für DIN EN ISO 9241 (DIN 2008, S. 18)

Die Norm besteht aus mehreren Teilen, wobei sich die Abschnitte 11 bis 17 und 110 dem Thema Software-Ergonomie widmen. Die Richtlinie 11 liefert die Definition zur Gebrauchstauglichkeit und stellt daher das „Kernstück“ der Definition von Usability dar. Sie beinhaltet vier Leitsätze: Leicht zu erlernen, intuitiv zu benutzen, geringe Fehlerrate sowie Zufriedenheit sicherstellen.

Der ursprüngliche Abschnitt 9241.10 „Grundsätze der Dialoggestaltung“ wurde ebenfalls 2006 umbenannt und trägt gegenwärtig die Kennziffer 9241.110. Dieser Abschnitt ist für die Entwicklung gebrauchstauglicher Webseiten essentiell. Die Grundsätze beziehen sich auf die Benutzer-/Benutzerinnenschnittstelle (engl. *User Interface*). Die Benutzer-/Benutzerinnenschnittstelle beschreibt die Bestandteile eines interaktiven Systems, welches die Steuerelemente zur Verfügung stellt, die für die Nutzer und Nutzerinnen notwendig sind, um eine bestimmte Aufgabe zu erledigen oder ein bestimmtes Ziel zu erreichen. Die Anforderungen an eine ausgeprägte Usability sind diesem Abschnitt zugeordnet.

Die Usability setzt sich nach DIN EN ISO 9241-110 zusammen aus der Effektivität, der Effizienz und der Zufriedenstellung einer Webseite (vgl. Kap. 2.6.1). Eine Webseite ist so zu gestalten, dass die Nutzer und Nutzerinnen ihr Ziel (Effektivität) mit möglichst wenig Aufwand (Effizienz) erreichen, um ein für sie zufriedenstellendes Ergebnis (Zufriedenheit) zu erzielen. Die Attribute Effektivität, Effizienz und Zufriedenheit werden im Folgenden kurz auf der Grundlage der DIN EN ISO 9241-110 erläutert.

Effektivität beschreibt den Grad, ob und wie genau ein Nutzer/eine Nutzerin sein/ihr Ziel erreicht. In Zusammenhang stehen damit die Gestaltung der Menü-, Navigations- und Orientierungsmittel, welche dies zulassen müssen. Die Effektivität entscheidet darüber, wie viel kognitiver Aufwand aufgebracht werden muss, um das Angebot formal handhaben zu können und wie viele kognitive Ressourcen demnach für das tiefere Verständnis und ggf. dem Wissenserwerb und Lernen verfügbar bleiben. Indikatoren für die Messung der Effektivität sind:

- ob das Ziel erreicht wurde
- welche/wie viele Informationen aufgerufen wurden
- welche/wie viele relevante Informationen nicht genutzt wurden.

Effizienz beschreibt das Verhältnis zwischen dem Aufwand und dem Ergebnis. Es werden beispielsweise Anstrengung, Aufrechterhaltung der Motivation und Zeit berücksichtigt. Ein Indikator zur Messung der Effizienz kann demnach die Zeit sein, die zur Erreichung des Ziels benötigt wird (vgl. auch Niegemann, 2008, S. 421; Rogers, Preece & Sharp, 2011, S. 14).

Zufriedenheit entsteht, wenn die Erwartungen des Nutzers/der Nutzerin und die Arbeitsprozesse ohne Einschränkungen erfüllt werden können. Die Zufriedenheit betrifft jegliche Aspekte der Gebrauchstauglichkeit. Zufriedenheit wird

gängigerweise über Fragebögen oder *Thinking Aloud*-Methoden (vgl. Kap. 2.7) erhoben (vgl. auch Niegemann, 2008, S. 421).

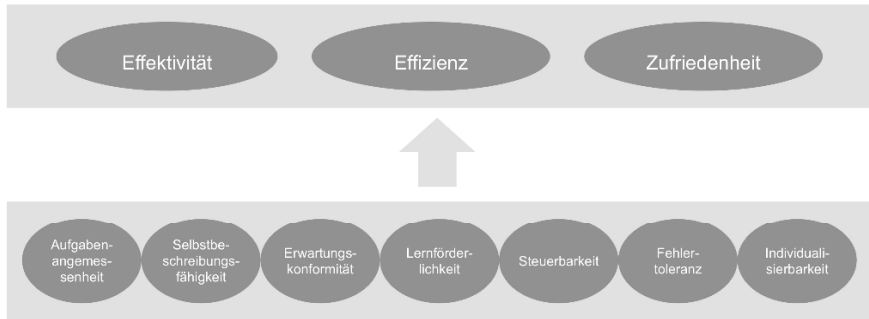


Abbildung 6: Grundsätze der Usability in Anlehnung an DIN EN ISO 9241-110 9241 (DIN 2008, S. 22)

Diese Attribute münden wiederum in sieben unterstützende Grundsätze der Dialoggestaltung (vgl. Abbildung 6): 1. Aufgabenangemessenheit, 2. Selbstbeschreibungsfähigkeit, 3. Erwartungskonformität, 4. Lernförderlichkeit, 5. Steuerbarkeit, 6. Fehlertoleranz und 7. Individualisierbarkeit (DIN Deutsches Institut für Normung e. V., 2008, S. 7):

1. „Ein interaktives System ist aufgabenangemessen, wenn es den Benutzer unterstützt, seine Arbeitsaufgabe zu erledigen, d. h., wenn Funktionalität und Dialog auf den charakteristischen Eigenschaften der Arbeitsaufgabe basieren, anstatt auf der zur Aufgabenerledigung eingesetzten Technologie“ (DIN Deutsches Institut für Normung e. V., 2008, S. 8).

Der Nutzer/die Nutzerin sollte auf einem direkten Weg das Ziel ohne zusätzliche belastende kognitive Anforderungen erreichen. Um diesen Anspruch zu erfüllen, sind im Vorfeld beispielsweise Aufgaben, Arbeits- und/oder Lernabläufe zu definieren, die einen potenziellen Nutzer/eine potenzielle Nutzerin ausführen möchte. Anschließend ist zu formulieren, wann eine Aufgabe erfüllt ist, um die Effektivität sicherzustellen. Letztlich ist der Weg zu definieren, der effizient – also mit geringem Einsatz von Zeit, Geduld, Gedächtnis- und Transferleistung – zu diesem Ziel führt.

2. *„Ein Dialog ist in dem Maße selbstbeschreibungsfähig, in dem für den Benutzer zu jeder Zeit offensichtlich ist, in welchem Dialog, an welcher Stelle im Dialog er sich befindet, welche Handlungen unternommen werden können und wie diese ausgeführt werden können“ (DIN Deutsches Institut für Normung e. V., 2008, S. 10).*

Dieser Definition nach muss sich der Benutzer/die Benutzerin zu jeder Zeit folgende drei Fragen beantworten können: Wo komme ich her? Wo bin ich? Wo kann ich von hier aus hin? Für die Beantwortung dieser Fragen ist die Besucherin/der Besucher in folgenden Aspekten zu unterstützen: a) Orientierung, b) Antizipierbarkeit, c) Feedback und d) Hilfe.

- a) Den Nutzer und Nutzerinnen müssen auf jeder Seite Orientierungspunkte angeboten werden, um zu erkennen, wo er/sie sich befindet und wie weit er/sie vom Ziel entfernt ist.
- b) Die Nutzer und Nutzerinnen müssen erkennen (antizipieren) können, wo die Navigationselemente hinführen. Erst damit wird die Seite steuer- und beherrschbar.
- c) Um den Nutzern und Nutzerinnen das Gefühl des Vertrauens zu geben, hat das System eine Rückmeldung darüber zu geben, ob Aktionen erfolgreich durchgeführt worden sind.
- d) Jedes System sollte Hilfestellungen anbieten, um auch unerfahrenen Nutzern und Nutzerinnen die Nutzung zu ermöglichen, insbesondere wenn die Seite und die auszuführenden Aktionen komplex sind.

Insgesamt ist das Ziel der Selbstbeschreibungsfähigkeit, dass Nutzer und Nutzerinnen in der Lage sind, die Software, die Internetseite etc. intuitiv richtig zu nutzen.

3. *„Ein Dialog ist erwartungskonform, wenn er den aus dem Nutzungskontext heraus vorhersehbaren Benutzerbelangen sowie allgemein anerkannten Konventionen entspricht“ (DIN Deutsches Institut für Normung e. V., 2008, S. 11).*

Das menschliche Verhalten ist geprägt durch erlernte Verhaltensmuster und Gewohnheiten. Je schneller Muster erkannt werden, desto schneller kann das Gehirn auch Zusammenhänge erfassen. Wenn Navigationselemente mit den gleichen Funktionen von Seite zu Seite variieren, können Nutzer und Nutzerin-

nen kein Muster speichern, welches ihnen die Navigation auf den folgenden Seiten erleichtern würde. Somit stünden auch weniger kognitive Ressourcen für den Inhalt zur Verfügung (vgl. Kap. 2.4 zur CLT). Je mehr Erfahrungen Nutzer und Nutzerinnen mit Software und Internetseiten haben, desto größer und präziser werden auch ihre Ansprüche bezüglich der Konformität mit ihren Erwartungen.

4. *„Ein Dialog ist lernförderlich, wenn er den Benutzer beim Erlernen der Nutzung des interaktiven Systems unterstützt und anleitet“ (DIN Deutsches Institut für Normung e. V., 2008, S. 12).*

Den Nutzern und Nutzerinnen ist die schnelle Aneignung von relevantem Wissen und Fertigkeiten zur Nutzung des Systems zu ermöglichen. Das Maß der Unterstützungsnotwendigkeit ist wiederum abhängig von den bisherigen Erfahrungen und gespeicherten mentalen Modellen. Es ist nicht möglich, jegliche individuelle Ausprägungen zu berücksichtigen, jedoch gilt, je seltener ein System genutzt oder eine Seite besucht wird, desto weniger ist ihm/ihr zuzumuten, die Bedienung zu erlernen (Rampl, 2007). Die Beanspruchung zusätzlicher kognitiver Ressourcen führt im Idealfall lediglich zur Verlangsamung der Interaktion, im schlimmsten Fall jedoch zur Unfähigkeit, den Dialog erfolgreich zu beenden. So komplex der Dialog auch sein mag, die Nutzer und Nutzerinnen verfügen über mentale Modelle, die ihnen mit geringer Transferleistung die Bedienung eines auf diese Muster angelegten Systems erlauben (Rampl, 2007). Nur in seltenen Fällen erfordert die Nutzung eines Systems das Erlernen neuer Navigationsarten. Das System sollte allerdings den Anspruch an einen logischen Aufbau und logische Abläufe sowie die Einbindung von Hilfeseiten/Hilfefunktion erfüllen. Die Lernförderlichkeit bestimmt somit die Hürde, die von den Nutzern und Nutzerinnen überwunden werden muss, um mit dem System zu arbeiten und ist daher ein wesentlicher Aspekt der Usability. Insbesondere ist sie von Bedeutung, wenn keine Schulungsmaßnahmen für die Nutzung vorgesehen sind bzw. das System eigenständig genutzt werden soll. Ein Maß zur Bestimmung der Lernförderlichkeit ist die Zeit, die Nutzer und Nutzerinnen brauchen, um ein bestimmtes Niveau im Umgang mit dem System zu erreichen (Niegemann, 2008, S. 424).

5. *„Ein Dialog ist steuerbar, wenn der Benutzer in der Lage ist, den Dialogablauf zu starten sowie seine Richtung und Geschwindigkeit zu beeinflussen, bis das Ziel erreicht ist“ (DIN*

Deutsches Institut für Normung e. V., 2008, S. 13).

Dazu zählen die Möglichkeiten, Medien zu nutzen oder auszuschalten, Alternativen zum Navigieren, einer Sicherstellung der korrekten Funktionsweise der Zurück-Schaltfläche, jederzeit die Startseite erreichen zu können sowie abubrechen. in Bezug auf Lernprogramme bedeutet Steuerbarkeit, dass Lernende jederzeit den Prozess unterbrechen und zu einem späteren Zeitpunkt fortsetzen können.

6. *„Ein Dialog ist fehlertolerant, wenn das beabsichtigte Arbeitsergebnis trotz erkennbar fehlerhafter Eingaben entweder mit keinem oder mit minimalem Korrekturaufwand seitens des Benutzers erreicht werden kann“ (DIN Deutsches Institut für Normung e. V., 2008, S. 14).*

Einerseits soll das System selbst wenig bzw. keine Fehler aufweisen, andererseits soll seitens des Systems verhindert werden, dass der Nutzer/die Nutzerin überhaupt erst Fehler macht. Falls dennoch Fehler gemacht werden, sollte eine schnelle Korrektur möglich sein. Fehler sind in diesem Fall Aktionen, die nicht zum gewünschten Erfolg führen (Niegemann, 2008, S. 423). Die Fehlerrate wird über das Aufsummieren der unerwünschten Systemreaktionen gemessen und kann zur Verarbeitung in folgende Kategorien eingeteilt werden: Vermeidbare Fehler (z. B. Programmierfehler oder Fehler, die mit Hilfe einer sorgfältigen Evaluation hätten verhindert werden können), bekannte, nicht vermeidbare Fehler (beispielsweise das Vertippen mit der Tastatur, das unbeabsichtigte Abschieken eines Formulars) und nicht antizipierbare Fehler (Fehler, die aufgrund unerwarteten Besucherverhaltens passieren). Bei der Entwicklung von Systemen sind schlussfolgernd Fehler, soweit möglich und antizipierbar, im Vorfeld zu vermeiden sowie bei nicht vermeidbaren Fehlern Wege aufzuzeigen, die den Nutzern und Nutzerinnen eine selbstständige Behebung erlauben.

7. *„Ein Dialog ist individualisierbar, wenn Benutzer die Mensch-System-Interaktion und die Darstellung von Informationen ändern können, um diese an ihre individuellen Fähigkeiten und Bedürfnisse anzupassen“ (DIN Deutsches Institut für Normung e. V., 2008, S. 15).*

Das Angebot soll sich an das individuelle Vorwissen und die Fähigkeiten anpassen lassen können. Bezüglich Webseiten sollte es möglich sein, individuelle Einstellungen vornehmen zu können, um einer heterogenen Nutzer-/Nutzerinnengruppe die Nutzung zu ermöglichen. Gängige Anpassungsmöglichkeiten beziehen sich auf die Sprache, die Schriftgröße und Vergrößerungsfunktion. Auch Filterfunktionen sind ein Merkmal von Individualisierung. Diese werden z. B. bei Einstufungstests verwendet, um anhand des Kenntnisstandes die optimale Aufgabenauswahl bereit zu stellen. Somit müssen nicht alle Inhalte genutzt bzw. bearbeitet werden, sondern nur diejenigen, die an den Kenntnisstand angepasst sind (Niegemann, 2008, S. 423).

Die Grundsätze stellen die wesentlichen Qualitätskriterien für die Gebrauchstauglichkeit bei der Entwicklung von Webseiten dar. Doch sind diese aufgrund ihres Abstraktionsgrades nicht immer einfach zu interpretieren und umzusetzen (vgl. auch Defizite allgemeiner Usability-Richtlinien in diesem Kap.). Die dargestellten Grundsätze sind jedoch nicht alle auf die computerbasierte Diagnostik zu übertragen. So kann die Steuerbarkeit auch absichtlich verhindert werden, um valide Ergebnisse zu erzielen: Während der Nutzung einer computerbasierten Diagnostik können ggf. Lerneffekte entstehen, so dass Nutzer und Nutzerinnen beispielsweise eine fehlerhafte Performanz korrigieren könnten. Auch die Individualisierbarkeit kann unter Umständen zu einem Reliabilitäts- und Validitätsproblem führen – je nach Abhängigkeit des Individualisierungsgrades und der Passung zwischen den Auswahlmöglichkeiten der computerbasierten Diagnostik und den Bedürfnissen oder dem Vorwissen der Nutzer und Nutzerinnen.

Neben der DIN Norm existieren weitere Qualitätskriterien, die je nach Nutzungskontext herangezogen werden können. Nielsen führt fünf Kriterien an, welche die Qualität der Usability definieren (Nielsen, 2012), die sich zum Teil mit den Normen der DIN überschneiden⁴⁸:

1. Erlernbarkeit (engl. *Learnability*): Wie einfach ist es für einen Nutzer/eine Nutzerin, bei der ersten Interaktion mit dem Interface, einfache Aufgaben auszuführen?
2. Effizienz (engl. *Efficiency*): Wie schnell können Nutzer und Nutzerinnen Aufgaben ausführen, sobald sie sich an das Design gewöhnt bzw. das Design „gelernt“ haben?
3. Erinnerbarkeit (engl. *Memorability*): Wie schnell können Nutzer und Nutzerinnen das erlernte Design reaktivieren, wenn sie über einen längeren Zeitraum die Seite bzw. die Umgebung nicht besucht haben?

⁴⁸ Auch die englischen Begriff finden im deutschen Sprachraum Verwendung, daher wird für jeden deutschen Begriff auch immer die englische Übersetzung aufgeführt.

4. Fehlertoleranz (engl. *Errors*): Wie viele Fehler können Nutzer und Nutzerinnen machen, wie schwerwiegend sind diese und wie einfach können sie diese Fehler beheben?
5. Zufriedenheit (engl. *Satisfaction*): Wie angenehm ist es die Seite zu nutzen?

In diese fünf Qualitätskriterien sind die Attribute Effizienz und Zufriedenheit der DIN EN ISO 9241-110 eingeschlossen, die Effizienz wird jedoch nicht explizit erwähnt.

Ein weiteres Qualitätskriterium, das Nielsen zusätzlich anführt aber nicht direkt der Qualität der Usability zuordnet, ist die Nützlichkeit, welche sich auf die Funktionalität bezieht. Hierbei geht es darum, in welchem Maß das Design die Absicht des Nutzers/der Nutzerin unterstützt. Nach Nielsen sind Usability und Utility gleich wichtig (vgl. auch Kap. 2.6.1 zur Begriffsan- und -einordnung).

Fisk u. a. (2009) formulieren Usability-Anforderungen für ältere Personen, die sich tendenziell durch eine niedrigere Computerkompetenz und eingeschränkte physische Leistungen (Sehfähigkeit und motorische Fähigkeiten) auszeichnen. Sie führen acht Prinzipien an, die sich ebenfalls zum Teil mit denen der DIN EN ISO 9241-110 und denen von Nielsen überschneiden (Fisk u. a., 2009, S. 75):

1. *Kompatibilität*: Das System sollte kompatibel mit den Erwartungen der Nutzer und Nutzerinnen sein.
2. *Konsistenz*: Die Platzierung von Symbolen und Schaltflächen sollte über die Screens hinweg gleich angeordnet sein und die gleichen Funktionen erfüllen.
3. *Fehlertoleranz*: Fehler von Nutzer und Nutzerinnen sollten von ihnen selbst leicht wieder behoben werden können.
4. *Feedback*: Die Resultate von Funktionen sollen klar sein und es sollte angezeigt werden, an welcher Stelle (der Bearbeitung) sie sich befinden.
5. *Individualisierbarkeit*: Nutzer und Nutzerinnen sollen die Möglichkeit haben, das System nach ihren Bedürfnissen anzupassen.
6. *Gedächtnis*: Das Gedächtnis der Nutzer und Nutzerinnen sollte nicht überladen sein und Unterstützungsfunktionen sollen angeboten werden.
7. *Struktur*: Die Interfaces sollen eine Struktur aufweisen, die den Nutzern und Nutzerinnen die Handhabung erleichtert.
8. *Workload*: Der Workload sollte so gering wie möglich gehalten werden, z. B. durch Hervorhebung von wichtigen Informationen.

Diese Guidelines sind zunächst abstrakt formuliert und werden hinsichtlich physischer Charakteristiken, Navigation, Informationsorganisation sowie konzeptioneller Ansprüche konkretisiert. Sie finden sich z. B. auch in der DIN EN ISO

9241-110 und in denen des United States Dept of Health and Human Services (2006) wieder. Allerdings liegt hier der Fokus darauf, spezielle Charakteristiken – wie z. B. eine Sehschwäche – zu berücksichtigen, indem beispielsweise die Schriftgröße anpassbar ist.

Molich und Nielsen (1990) haben versucht, die Komplexität der Regelbasis von Usability-Guidelines zu reduzieren und grundsätzliche Usability-Prinzipien als neun Heuristiken formuliert (Molich & Nielsen, 1990, S. 339):

1. Einfacher und natürlicher Dialog (engl. *Simple and natural dialogue*): Der Dialog am Bildschirm soll keine irrelevanten oder wenig nützlichen Informationen enthalten, denn jede irrelevante Information konkurriert mit den unverzichtbaren Aspekten des Dialogs. Die Informationen sollten in einer natürlichen und logischen Abfolge erscheinen.
2. Sprich die Sprache des Nutzers/der Nutzerin (engl. *Speak the user's language*): Der Dialog ist in verständlichen Wörtern und Sätzen zu erfolgen, die an die antizipierte Sprache potenzieller Nutzer und Nutzerinnen angepasst ist, anstatt einen systemorientierten Ausdruck bzw. Fachsprache der Computertechnologie zu verwenden.
3. Minimiere die Gedächtnisbelastung der Nutzerin/des Nutzers (engl. *Minimize user memory load*): Das Arbeitsgedächtnis ist begrenzt (vgl. hierzu auch Kap. 2.4 zur CLT). Somit sollte von den Nutzern und Nutzerinnen nicht gefordert werden, sich an Informationen aus vorherigen Dialogen zu erinnern. Instruktionen zur Nutzung des Systems sollten klar sichtbar oder abrufbar an solchen Stellen sein, an denen sie auch gebraucht werden. Zudem sind Instruktionen in leicht verständlichem Ausdruck zu formulieren.
4. Sei konsistent (engl. *Be consistent*): Nutzer und Nutzerinnen sollten sich nicht fragen müssen, ob unterschiedliche Wörter, Situationen oder mögliche Aktionen das gleiche bedeuten. Eine bestimmte immer wiederkehrende Aktion sollte auch immer wieder durch die gleiche Aktion erreichbar sein (z. B. durch Anklicken einer bestimmten Schaltfläche).
5. Biete Feedbackfunktionen an (engl. *Provide feedback*): Das System sollte der Nutzerin/die Nutzerin permanent darüber informieren, wo er/sie sich befindet und welche Aktionen ausgeführt werden.
6. Platziere Möglichkeiten zum Verlassen des Systems klar und deutlich (engl. *Provide clearly marked exits*): Nutzer und Nutzerinnen sollten nicht in Bereichen „gefangen“ gehalten werden, indem keine Möglichkeit angeboten wird, das System oder die Seite zu verlassen. Es sollte immer die Möglichkeit eines „Notausgangs“ geben, der das schnelle Verlassen ohne das Durchlaufen eines längeren Dialogs ermöglicht.

7. Ermögliche die Nutzung von Shortcuts (engl. *Provide shortcuts*): Die Features, die ein System leicht erlernbar machen – wenig Eingabefelder und wortreiche Dialoge – sind oft beschwerlich für erfahrene Nutzer und Nutzerinnen. Intelligente Shortcuts sollten so in ein System integriert sein, dass das System sowohl erfahrene als auch unerfahrene Nutzer und Nutzerinnen anspricht.
8. Gute Fehlermeldungen (engl. *Good error messages*): Gute Fehlermeldungen sind defensiv, präzise und konstruktiv. Sie geben als Ursache immer das System an und kritisieren nicht den Nutzer/die Nutzerin.
9. Fehlerprävention (engl. *Error Prevention*): Noch besser als gute Fehlermeldungen ist es, Fehler und Probleme durch umsichtiges Design zu vermeiden.

Diese Heuristiken scheinen auf den ersten Blick offensichtlich zu sein, doch bei ihrer Überprüfung wird schnell deutlich, dass die Umsetzung nicht trivial ist. Molich und Nielsen (1990) stellten fest, dass Programmierer und Programmierinnen sowie Entwickler und Entwicklerinnen Schwierigkeiten in der Identifizierung potenzieller Probleme eines simplen Computerdialogs haben. Aus einer Evaluation eines Computerdialogs folgerten sie, dass überhaupt erst mal das Risiko fehlerhaften Designs von den Evaluierenden realisiert werden müsse. Darauf folgend sind Dialogprinzipien, wie die von ihnen formulierten, zu berücksichtigen um letztlich die Usability von vielen potenziellen Nutzern und Nutzerinnen überprüfen zu lassen. Werden Computersysteme von Personen entwickelt, die grundsätzliche Prinzipien der Dialoggestaltung verstehen und berücksichtigen, erreichen sie auch eine stärker ausgeprägte Usability. Die Ergebnisse der Evaluation weisen darauf hin, dass die Prinzipien weder selbstverständlich noch intuitiv bei der Entwicklung berücksichtigt werden (Molich & Nielsen, 1990, S. 342).

Stark daran angelehnt sind die „8 golden Rules“ nach Shneiderman und Plaisant (2005). Erweiternd werden von ihnen die Heuristiken mit *Design dialogue to yield closure*, *Permit easy reversal of actions* und *Support internal locus of control* angeführt (Shneiderman & Plaisant, 2005, S. 75). Die restlichen vier Regeln decken die neun Prinzipien von Molich und Nielsen ab. *Design dialogue to yield closure* beschreibt die Anforderung, Handlungssequenzen zu gruppieren in Anfang, Mitte und Ende. Eine Rückmeldung am Ende einer Handlungssequenz gibt dem Nutzer/der Nutzerin das Gefühl der Zufriedenheit, etwas abgeschlossen zu haben, ein Gefühl der Entlastung bzw. der Ablösung durch eine darauf folgende Handlungssequenz. *Permit easy reversal of actions* erhebt den Anspruch, dass Aktionen rückgängig zu machen sind. So wird Angst vor Fehlern aus der Perspektive der Nutzenden verringert und unterstützt das Ausprobieren von ggf. noch nicht bekannten Optionen. Die Heuristik *Support internal locus of*

control fordert, dass die Nutzer und Nutzerinnen überzeugt sind, die Kontrolle über ihre Handlungen zu haben. Dies beinhaltet auch die Möglichkeit, notwendige Informationen abrufen und notwendige Handlungen ausführen zu können. Im Zuge der technischen Entwicklung, der Standardisierung (inklusive der Richtlinien für die Barrierefreiheit – vgl. Kap. 2.6.4.2) und kulturellen Entwicklung können Heuristiken überholt, nicht mehr zutreffend sein und an Relevanz verlieren.

So haben Sarodnick und Brau (2011) haben 21 Jahre nach Nielsen und Molich 12 alternative Heuristiken zu den Heuristiken von Molich und Nielsen (1990) formuliert. Diese berücksichtigen sowohl die Prinzipien der bisher geläufigen als auch neue technische und kulturelle Entwicklungen. Erweitert werden die neun Heuristiken von Nielsen und Molich um die Heuristiken der *Individualisierbarkeit*, *Joy of use* und *Interkulturelle Aspekte* (Sarodnick & Brau, 2011, S. 149 f). *Individualisierbarkeit* bedeutet, das Dialogsystem sollte sich den individuellen Präferenzen anpassen lassen, so lange es der Effektivität, Effizienz und Zufriedenstellung dient und nicht im Widerspruch zu notwendigen technischen oder sicherheitsrelevanten Begrenzungen steht. *Joy of Use* bezieht sich darauf, dass Arbeitsabläufe und grafische Gestaltung – bei notwendiger Konsistenz – Monotonie vermeiden und zeitgemäß wirken sollten. Zudem sind Metaphern adäquat auf den Nutzungskontext abzustimmen. *Interkulturelle Aspekte* sollten dahingehend berücksichtigt werden, dass das System auf einen definierten Nutzer-/Nutzerinnenkreis und dessen funktionale, organisatorische und nationale Kultur abgestimmt sein sollte.

In den vorgestellten Heuristiken teilweise „mitgedacht“ aber nicht explizit berücksichtigt sind Personen mit körperlichen und psychischen Einschränkungen. Die WCAG knüpfen an den bereits vorgestellten Heuristiken an, werden jedoch um Empfehlungen speziell für Personen mit Behinderungen erweitert.

2.6.4.2 Accessibility und die Web Content Accessibility Guidelines (WCAG)

In dem Kapitel zur computerbasierten Diagnostik wurden Beispiele für Diagnoseinstrumente genannt, die für einen bestimmten Anlass entwickelt wurden, z. B. Vorauswahl von potenziellen Bewerbenden oder Steigerung der Literalität. Doch auch alltägliche Handlungen werden zunehmend über das Internet vollzogen – angefangen vom Einkaufen bis hin zur Anmeldung bei Behörden. Auch Personen mit Behinderungen nutzen zunehmend das Internet. Personen mit Behinderungen umfassen Personen mit visuellen, auditiven, motorischen, sprachlichen, kognitiven, Sprach-, Lern- und neurologischen Einschränkungen. Wie in den Kapiteln zu physiologischen und kognitiven Voraussetzungen erläutert, können Einschränkungen in diesen Bereichen einen großen Einfluss darauf haben, wie ein Interface wahrgenommen wird und insbesondere, ob eine computerbasierte

Diagnostik valide Ergebnisse liefert. Somit wird ein barrierefreier Zugang zu Internetseiten immer wichtiger.

Die Barrierefreiheit ist, wie auch die Ansprüche an die Usability in einer ISO-Norm (40500:2012) (International Organization for Standardization & International Electrotechnical Commission, 2012) geregelt, den sog. *Web Content Accessibility Guidelines* (WCAG) 2.0 (vgl. auch Caldwell u. a., 2008). Diese sind auf die Zugänglichkeit interaktiver Systeme (einschließlich Software) bezogen und stellen eine Grundlage dar, Zugangsbarrieren für Personen mit physischen und psychischen Einschränkungen abzubauen. Diese beziehen sich auf Interaktionen im Internet und basieren auf den vier Prinzipien (1) wahrnehmbar, (2) bedienbar, (3) verständlich und (4) robust. Darunter aufgeschlüsselt sind zwölf Richtlinien. Die Prinzipien und die dazugehörigen Richtlinien werden im Folgenden auf der Grundlage der Angabe des World Wide Web Consortiums (W3C) kurz dargestellt (Caldwell u. a., 2008):

Prinzip 1: Wahrnehmbar – Informationen und Bestandteile der Benutzer-/Benutzerinnenschnittstelle müssen dem Benutzer/der Benutzerin so präsentiert werden, dass sie diese wahrnehmen können. Hierzu zählen Richtlinien zu Textalternativen, zeitbasierten Medien (wie z. B. Audio und Video), Anpassbarkeit und Unterscheidbarkeit (beispielsweise bezüglich Farben und Kontraste).

Prinzip 2: Bedienbar – Bestandteile der Benutzer-/Benutzerinnenschnittstelle und Navigation müssen bedienbar sein. Richtlinien zu diesem Prinzip beziehen sich auf die Zugänglichkeit durch die Tastatur, ausreichend Zeit, Vermeidbarkeit von (epileptischen) Anfällen (z. B. dürfen dargestellte Inhalte nicht mehr als dreimal in kurzen Abständen blitzen) und Navigierbarkeit.

Prinzip 3: Verständlich – Informationen und Bedienung der Benutzer-/Benutzerinnenschnittstelle müssen verständlich sein. Hierzu zählen Lesbarkeit, Vorhersehbarkeit und Hilfestellung bei der Eingabe.

Prinzip 4: Robust – Inhalte müssen robust genug sein, damit sie zuverlässig von einer großen Auswahl an Benutzer und Benutzerinnen, einschließlich assistierender Techniken, interpretiert werden können. Dem Prinzip der Robustheit ist die Richtlinie zur Kompatibilität zugeordnet. Die Richtlinien geben die inhaltlichen Ziele vor, auf die Autoren hinarbeiten sollten, um Inhalte für die Nutzer und Nutzerinnen mit verschiedenen Behinderungen barrierefreier zu gestalten. Auch wenn die Richtlinien zur barrierefreien Gestaltung ein weiteres Feld an Bedürfnissen abdecken, können nicht alle Arten, Ausprägungen und Kombinationen von Behinderungen adressiert werden. Die Richtlinien verbessern die Nutzbarkeit explizit z. B. für ältere Personen mit sich altersbedingt ändernden Fähigkeiten. Sie verbessern damit häufig aber auch die Gebrauchstauglichkeit für Nutzer und Nutzerinnen im Allgemeinen (Caldwell u. a., 2008). Werden die

Ansprüche an einen barrierefreien Zugang berücksichtigt, steigt durch die klaren und einfachen Strukturen die Gebrauchstauglichkeit, die Kompatibilität wird verbessert und die Ladezeiten werden verkürzt. Das Thema *Accessibility* (engl. für Barrierefreiheit, Erreichbarkeit, Zugänglichkeit) im Kontext computerbasierter Diagnostik zu berücksichtigen, da diese auch online und ohne weitere Unterstützung stattfinden kann. Zudem ist die Zielgruppe der in diesem Forschungsprojekt untersuchten online und anonym zugänglichen Online-Plattform otu.lea tendenziell wenig gewohnt mit dem Computer zu arbeiten und profitiert vermutlich ebenfalls von einer barrierefreien Gestaltung. Die Zielgruppe von funktionalen Analphabeten und Analphabetinnen ist allerdings nicht mit der Zielgruppe der WCAG gleichzusetzen. Es kann zwar davon ausgegangen werden, dass einige funktionale Analphabeten und Analphabetinnen neben literalen Defiziten gleichzeitig Lernbehinderungen aufweisen (empirische Untersuchungen gibt es darüber bisher nicht, diese Vermutung gründet sich auf Erfahrungsberichten von Kursleitenden in der Alphabetisierung), jedoch ist diese Gruppe nicht durch Lernbehinderung oder physische Einschränkungen zu charakterisieren. Bei der Entwicklung einer Internetplattform für tendenziell computerungewohnte ist es daher zwar ratsam, auf die barrierefreie Nutzung zu achten – insbesondere im Kontext von onlinebasierten Kompetenzmessungen, da der Einfluss der ICT-Literacy auf das Testergebnis so weit wie möglich auszuschließen ist, um eine möglichst hohe inhaltliche Validität zu erzielen – jedoch ist es nicht notwendig und auch nicht ratsam, jegliche Richtlinien der WCAG zu berücksichtigen. Es ist zu vermuten, dass Anpassungsmöglichkeiten, die speziell für Personen mit Einschränkungen entwickelt wurden, auf eine geringe Akzeptanz stoßen und ggf. – aufgrund der empfundenen Gleichsetzung mit physisch oder psychisch eingeschränkten Personen – zum Abbruch führen würde.

Im Vergleich zu den Usability-Heuristiken wird hier deutlich, dass der Fokus stark auf die Nutzer und Nutzerinnen gerichtet ist und mit diesen Richtlinien der Anspruch erhoben wird, möglichst viele Personen mit unterschiedlichen Voraussetzungen den Zugang und die Nutzung zu ermöglichen. Gleichzeitig hat das Erfüllen bzw. Nicht-Erfüllen dieser Richtlinien einen Einfluss auf die Akzeptanz der potenziellen Nutzer und Nutzerinnen. Die DIN EN ISO 9241-110 versucht ebenfalls dem Anspruch gerecht zu werden, vielen Nutzern und Nutzerinnen den Zugang zu ermöglichen und schließt in ihren Richtlinien die barrierefreie Nutzung von Software zu einem großen Teil ein. Die WCAG liefern zusätzlich zu den Richtlinien ausführliche Hinweise, wie die Anforderungen inhaltlich und technisch erfüllt werden können. Es wird dabei allerdings weniger berücksichtigt, wie bei dem Entwicklungsprozess vorzugehen ist, um diesen Ansprüchen gerecht zu werden, zudem sind die Richtlinien und Prinzipien nicht explizit auf computerbasierte Diagnostik bezogen. Da aber auch in Institutionen

wie beispielsweise Schulen und Universitäten vermehrt computerbasierte Tests durchgeführt werden, bekommt die Barrierefreiheit auch hinsichtlich der computerbasierten Diagnostik eine stärkere Relevanz. Wird bei Tests in Universitäten und Schulen nicht auf eine barrierefreie Gestaltung geachtet, hat das womöglich Konsequenzen für die Schüler und Schülerinnen oder Studierende, da es ihnen unter Umständen nicht oder nur bedingt möglich ist aufgrund von Behinderung(en) an dem Test teilzunehmen und diesen erfolgreich abzuschließen. In England fand diese Überlegung ihren Anfang in dem *Disability Discrimination Act* (1995), in den USA mündeten die Bemühungen u. a. in dem *No Child Left Behind Act* 2001. Diese Anforderungen sind aber ebenfalls nicht auf elektronische Tests fokussiert. Erst 2002 wurde von der IMS Accessibility Project Group (IMS Global Learning Consortium) eine internationale Richtlinie für die Entwicklung von Lernapplikationen entwickelt (Crisp, 2007, S. 168). Diese Richtlinien und Standards führen allerdings bei Berücksichtigung nicht auch automatisch zu einem hohen Qualitätsstandard oder einer hohen Gebrauchstauglichkeit. Diese Lücke versucht Ball (2009) zu schließen, indem er den Fokus auf den Entwicklungsprozess barrierefreier E-Assessments richtet. Er formuliert vier Prinzipien für Entwickler und Entwicklerinnen, die es bei der Entwicklung eines E-Assessments zu berücksichtigen gilt (Ball, 2009):

- a) Prinzip des Antizipierens: Möglichst alle möglichen personalen Voraussetzungen für die Zugänglichkeit von E-Assessments müssen in dem Entwicklungsprozess berücksichtigt werden.
- b) Prinzip der Begründeten Anpassung: In vielen Entwicklungsprozessen begrenzt das Budget auch die Zeitspanne, in der Entwicklungen umgesetzt werden. Das hat oft eine reduzierte Umsetzung von Accessibility-Ansprüchen zur Folge. Es ist darauf zu achten, dass auf der Grundlage der materiellen Ressourcen klare Ziele formuliert werden.
- c) Prinzip der permanenten technischen Entwicklung: Im Zuge der technischen Entwicklungen werden auch neue Möglichkeiten für die Verbesserung der Zugänglichkeit entstehen. Daher sind die E-Assessments hinsichtlich ihrer Zugänglichkeit wiederholt zu überprüfen und die Usability ist entsprechend anzupassen.
- d) Prinzip der gemeinsamen Verantwortung: Die entwickelnde Institution hat eine klare Antidiskriminierungspolitik zu verfolgen. Zudem sollten sich alle Beteiligten aktiv für die Einhaltung dieser vier Prinzipien aussprechen.

Ball schlägt zudem praktische Schritte für die Entwicklung eines barrierefreien E-Assessments vor – angefangen von der Zusammenstellung des Entwickler-/Entwicklerinnen-Teams bis hin zum Akzeptanztest (Ball, 2009).

Um diesen Prinzipien und Anforderungen gerecht zu werden, bietet die momentane technische Entwicklung zahlreiche Möglichkeiten. Auf der technischen Ebene ist die Einbindung von Audio, Video und automatisierten Feedbackfunktionen möglich. Auf der Design-Ebene können Materialien „authentisch“ dargestellt werden. Die Diagnostik wird so mit multimedialen Elementen angereichert (vgl. Kap. 2.2.3 zum Thema Rich E-Assessment) und kann damit förderdiagnostischen Ansprüchen (vgl. Kap. 2.1.2) genügen.

2.6.4.3 Heuristiken – Grenzen der Anwendbarkeit

Auch wenn sich die Anwendung solcher Guidelines und Heuristiken etabliert hat, impliziert die Anwendung diverser Richtlinien nicht gleichzeitig, dass ein Interface auch gebrauchstauglich ist. Im Folgenden werden die Grenzen der Anwendbarkeit von Heuristiken (die auch auf Guidelines übertragbar sind) aufgezeigt.

Heuristiken sind immer kontextabhängig zu betrachten und anzuwenden. Vor der Anwendung ist zu prüfen, welche Heuristiken für den jeweiligen Zweck und die jeweilige Zielgruppe sinnvoll erscheinen und ob sie in Teilaspekten modifiziert werden müssen. Die Schwäche von Heuristiken liegt darin, dass sie meist nicht für eine spezielle Zielgruppe und Anwendungsumgebungen formuliert sind und meist unklar ist, wo die Grenzen der Interpretier- und Anwendbarkeit verlaufen (Bevan u. a., 1991, S. 654), woraus sich insbesondere drei Defizitbereiche von Guidelines ableiten lassen (Nielsen & Mack, 1994):

- Die Aussagekraft von Heuristiken hängt stark von dem Grad der Generalisierung ab. Je genereller eine Heuristik formuliert ist, desto weitläufiger sind auch Anwendungsfeld und Interpretationsspielraum. Um eine standardisierte Guideline zu interpretieren und anzuwenden, sind das Objekt, der Anwendungskontext, die potenzielle Zielgruppe sowie deren angenommene Eigenschaften hinsichtlich Präferenzen, Performanz und Lesefähigkeit zu definieren.
- Generelle Heuristiken können nicht „per se“ angewendet werden. Für die Anwendung bedarf es einer konkreten Definition des Nutzungskontextes: Die experimentellen Bedingungen, unter denen Richtlinien entwickelt und erprobt worden, sind nicht immer transparent. Somit können Heuristiken suggerieren, auf viele Anwendungskontexte übertragbar zu sein, ohne jedoch im Vorfeld einer Überprüfung unterzogen worden zu sein. Somit sind mit der Formulierung weitere, bisher nicht berücksichtigte oder getestete Anwendungskontexte ausgeschlossen, ohne dass die Grenzen abgesteckt werden. Falls nicht erwähnt wird, wie und wo die Heuristiken anzuwenden sind, bergen generell formulierte Heuristiken die Gefahr einer Missinterpretation.

- Das in den Richtlinien verwendete Vokabular stammt meist aus unterschiedlichen Disziplinen. Um eine erfolgreiche Anwendung zu gewährleisten und Fehlinterpretationen zu vermeiden, sind erfahrene Entwickler und Entwicklerinnen (aus möglichst unterschiedlichen Disziplinen) notwendig.

Wie auch die einzelnen Heuristiken immer kritisch hinsichtlich des Anwendungskontexts zu reflektieren sind, gilt auch für die eben genannten Defizite die Kontextabhängigkeit. Dennoch sind Richtlinien und die Verwendung von Checklisten in vielen Fällen unabdingbar, um formalen Standards gerecht zu werden. Die Heuristiken und Heuristiken sind je nach Domänenspezifität und Evaluationsgegenstand zu erweitern oder anzupassen. Dies geschieht immer unter der Prämisse, dass dadurch die Effektivität der Evaluation qualitativ und/oder quantitativ verbessert werden kann (Sarodnick & Brau, 2011, S. 148). Um generelle Heuristiken um domänen- oder produktspezifische Anforderungen zu erweitern, existieren nach Sarodnick und Brau zwei Möglichkeiten: Entweder durch die Einführung zusätzlicher Heuristiken oder durch die Erweiterung bzw. Modifikation bereits bestehender Heuristiken (Sarodnick & Brau, 2011, S. 148). Beispielsweise kann es bei einer computerbasierten Diagnostik sinnvoll sein, die Heuristik bezüglich der Möglichkeit, Eingaben zu korrigieren, umzuformulieren. So sollte ggf. die Möglichkeit bestehen, Eingaben unmittelbar nach der Eingabe zu korrigieren aber nicht unbedingt, nachdem weitere Aufgaben bearbeitet wurden.

Problematisch ist allerdings, dass universal formulierte Heuristiken suggerieren, bei ihrer Anwendung auch universelle Usability zu erreichen. Für die Anwender und Anwenderinnen von Heuristiken besteht eine Schwierigkeit darin, Heuristiken einzuordnen hinsichtlich ihrer Relevanz für das spezifische System. Eine Studie von Mosier und Smith (1986) weist darauf hin, dass oft weniger als die Hälfte der genutzten Heuristiken aufgrund von zu starker Spezifität, mangelnder Spezifität, verwirrend oder nicht zutreffend angewandt werden (Hertzum, 2010, S. 571). Die Herausforderung einer universellen Usability besteht in der Nutzer-/Nutzerinnen-Diversität, Wissenslücken und Technik-Variationen (Shneiderman, 2000). Studien weisen zudem darauf hin, dass existierende Heuristiken und weitere universell formulierte Usability-Prinzipien tendenziell entweder zu stark generalisiert oder zu vielzählig sind, um akribisch angewandt zu werden (Hertzum & Jacobsen, 2003; Mosier & Smith, 1986; P. Reed u. a., 1999). Die Richtlinien und Anforderungen an die Usability werden auf der Grundlage der Annahme formuliert, dass die Interfaces die größtmögliche Bandbreite an menschlicher Heterogenität abdecken. Shneiderman (2000, S. 85) formuliert den Anspruch an eine universelle Usability, die für 90% der Haushalte zutrifft. Durch diesen Anspruch ist es aber nicht möglich, spezielle

Charakteristika von potenziellen Nutzer-/Nutzerinnengruppen zu berücksichtigen. Z. B. sind mit der Zunahme der Nutzung der Technologie, beispielsweise um Formulare für Behörden und Ämter auszufüllen, auch spezielle Charakteristiken von Personengruppen bei der Entwicklung zu berücksichtigen, die nicht „dem normalen Nutzer/der normalen Nutzerin“ entsprechen. Um dieser Herausforderung zu begegnen versucht Hertzum (2010) beispielsweise, verschiedene Arten der Usability zu benennen und voneinander abzugrenzen: universelle, situationelle, wahrgenommene, hedonistische, organisationale sowie kulturelle Usability. Diese Differenzierung hat sich bisher nicht durchgesetzt. Zudem deutet alleine die Benennung der unterschiedlichen Arten von Usability auf Teilaspekte der Usability hin, die in allgemein formulierten Heuristiken und Guidelines vertreten sind. Erfolgreich dahingegen sind die Ansätze, Usability-Guidelines für bestimmte Zielgruppen zu formulieren, z. B. für Ältere (vgl. z. B. Fisk u. a., 2009) und Personen mit kognitiven und physischen Einschränkungen. Damit ist erneut darauf hingewiesen, dass Heuristiken unter Berücksichtigung des Nutzungskontextes und der potenziellen Zielgruppe auszuwählen, ggf. zu modifizieren und anzuwenden sind.

2.6.4.4 Zusammenfassung der allgemeinen Usability-Anforderungen und der WCAG

Für die Darstellung der oben genannten Ansprüche an eine ausgeprägte Usability wurden die Bezeichnungen („Attribute“, „Prinzipien“, „Guidelines“ und „Heuristiken“) von den jeweiligen Autoren und Autorinnen übernommen. Werden diese miteinander verglichen, wird deutlich, dass viele Überschneidungen bestehen. Um eine Übersicht über die grundlegenden Ansprüche an die Usability zu erhalten, wurden die oben vorgestellten Heuristiken einer genaueren vergleichenden Betrachtung unterzogen, um die Komplexität und Vielzahl der Anforderungen zu reduzieren. Ähnliche Heuristiken, wie z. B. „Reduziere die Auslastung des Kurzzeitgedächtnisses“ (Nielsen & Molich, 1990; Shneiderman & Plaisant, 2005), „Aufgabenangemessenheit“ (DIN) und „Workload“ (Fisk u. a., 2009) wurden unter der jeweils höheren Abstraktionsebene – in diesem Beispiel „Aufgabenangemessenheit“ – zusammengefasst. Daraus ergeben sich folgende Heuristiken:

1. Erlernbarkeit (DIN, 2006; Nielsen, 2006; Sarodnick & Brau, 2011): Der Nutzer/Die Nutzerin wird beim Erlernen des Systems angeleitet und unterstützt.
2. Konsistenz (DIN, 2006; Fisk u. a., 2009; Nielsen & Molich, 1990; Shneiderman & Plaisant, 2005): Die Gestaltung, Platzierung und Funktion von Symbolen und Schaltflächen ist über die einzelnen Interfaces hinweg konsistent.

3. Fehlertoleranz (Caldwell u. a., 2008; DIN Deutsches Institut für Normung e. V., 2008; Fisk u. a., 2009; Nielsen, 2012; Nielsen & Molich, 1990; Shneiderman & Plaisant, 2005): Meldet das System einen Fehler, müssen Art und Handlungszusammenhang enthalten sein. Handlungen müssen reversibel sein oder der Nutzer/die Nutzerin muss über irreversible Handlungen informiert werden. Zudem sollte das System von vorne herein so gestaltet sein, dass möglichst wenig Fehler passieren können.
4. Feedback (Fisk u. a., 2009; Molich & Nielsen, 1990; Shneiderman & Plaisant, 2005): Die Konsequenzen von Aktionen sind absehbar. Das System sollte dem Nutzer/der Nutzerin Transparenz über den Bearbeitungsstand und die Stelle, an der er/sie sich im System befindet, gewährleisten. Zudem sollte dem Nutzer/der Nutzerin eine Rückmeldung darüber gegeben werden, ob eine Aktion erfolgreich abgeschlossen wurde.
5. Individualisierbarkeit (Caldwell u. a., 2008; DIN Deutsches Institut für Normung e. V., 2008; Fisk u. a., 2009; Sarodnick & Brau, 2011): Das System soll für die Bedürfnisse der jeweiligen Nutzer und Nutzerinnen anpassbar sein (beispielsweise sollen Einstellungsmöglichkeiten für die Schriftgröße, die Darstellung von Grafiken und die Ein-/Ausschaltung von Multimedia vorhanden sein).
6. Selbstbeschreibungsfähigkeit (DIN, 2006; Nielsen & Molich, 1990; Sarodnick & Brau, 2011): Nutzer und Nutzerinnen sind in der Lage, das System intuitiv richtig zu nutzen und jegliche Dialogschritte nachzuvollziehen.
7. Wahrnehmungssteuerung (Caldwell u. a., 2008; Fisk u. a., 2009; Nielsen & Molich, 1990; Sarodnick & Brau, 2011): Die Interfaces sollen eine Struktur aufweisen, die leicht erschließbar ist und den Nutzern und Nutzerinnen die Handhabung erleichtern. Dies kann beispielsweise durch die Berücksichtigung der Gestaltgesetze (vgl. Kap. 2.6.3.2) erreicht werden. Das Layout sollte minimalistisch gestaltet sein. Die Aufmerksamkeit sollte mit Hilfe gestalterischer Elemente, die ausreichend kontrastreich gestaltet sind, auf relevante Informationen gerichtet werden.
8. Aufgabenangemessenheit (DIN, 2006; Fisk u. a., 2009; Molich & Nielsen, 1990; Sarodnick & Brau, 2011; Shneiderman & Plaisant, 2005): Die Nutzer und Nutzerinnen werden darin unterstützt, ihre Aufgaben und Ziele effektiv und effizient zu erreichen, d. h. ohne das Arbeitsgedächtnis zusätzlich zu belasten.
9. Erwartungskonformität (DIN, 2006; Sarodnick & Brau, 2011): Die Gestaltung des Systems sollte möglichst den Erwartungen der Nutzer und

Nutzerinnen entsprechen und deren Erfahrungen sowie Vorwissen bezüglich der Systemnutzung berücksichtigen.

10. Steuerbarkeit (DIN, 2006; Nielsen & Molich, 1990; Sarodnick & Brau, 2011; Shneiderman & Plaisant, 2005): Nutzer und Nutzerinnen sollten selbst über Start, Ende, Unterbrechung und Geschwindigkeit der Nutzung bestimmen können.
11. Perspektivübernahme (Caldwell u. a., 2008; Nielsen & Molich, 1990; Shneiderman & Plaisant, 2005): Es sollte die Sprache möglichst vieler potenzieller Nutzer und Nutzerinnen gesprochen werden, so dass der Inhalt ein hohes Maß an Verständlichkeit aufweist. Im Vergleich zur Erwartungskonformität bezieht sich die Perspektivübernahme insbesondere auf die Verständlichkeit der Inhalte wohingegen die Erwartungskonformität auf Handlungen bezogen ist.
12. Prozessangemessenheit (Sarodnick & Brau, 2011): Das System sollte für die Erfüllung realer Aufgaben in typischen Einsatzfeldern optimiert sein.
13. Datensicherheit (Sarodnick & Brau, 2011): Das System sollte auch bei fehlerhaften Eingaben und auch unter hoher Ressourcenbelastung stabil und ohne Datenverlust arbeiten.
14. Shortcuts (Nielsen & Molich, 1990): Sowohl für erfahrene als auch unerfahrene Nutzer und Nutzerinnen sollte die Nutzung gängiger Shortcuts möglich sein.
15. Joy of use (Sarodnick & Brau, 2011): Ohne die Konsistenz zu beeinträchtigen, sollte Monotonie vermieden und zeitgemäße Gestaltungselemente eingebunden werden. Metaphern sind adäquat auf den Nutzungskontext abzustimmen.
16. Interkulturelle Aspekte (Sarodnick & Brau, 2011): Die kulturellen Hintergründe potenzieller Nutzer und Nutzerinnen sollten berücksichtigt und das System sollte darauf abgestimmt werden. Dies betrifft somit auch z. B. die Gestaltung des Storyboards und das Identifikationspotenzial mit dessen Charakteren.

Die Prinzipien der WCAG sind den vorgeschlagenen Heuristiken nicht eindeutig zuzuordnen. Diese überschneiden sich mitunter mehreren Heuristiken (so wäre z. B. das Prinzip „Verständlichkeit“ sowohl der Heuristik „Selbstbeschreibungsfähigkeit“ als auch der „Erwartungskonformität“ zuzuordnen, ebenso überschneiden sich Aspekte des Prinzips „Wahrnehmbar“ mit der Heuristik „Erwartungskonformität“), so dass die Prinzipien zwar in den Heuristiken mitunter berücksichtigt, nicht aber diesen zugeordnet sind.

Die Heuristiken stellen eine Zusammenfassung von Heuristiken für typische Nutzer und Nutzerinnen dar. In einem weiteren Schritt sind (vorerst auf theoreti-

scher Ebene) solche Heuristiken zu identifizieren, die ebenso für die in dieser Arbeit fokussierte Zielgruppe – funktionale Analphabeten und Analphabetinnen – geeignet sind.

2.6.5 Vorläufige Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen

Die Relevanz der Heuristiken sowohl für funktionale Analphabeten und Analphabetinnen als auch vor dem Hintergrund des Produkts – der computerbasierten Diagnostik – zu bewerten. Im Folgenden wird erläutert und begründet, welche Heuristiken für die Entwicklung von otu.lea berücksichtigt wurden.

Ausgehend von der eben dargestellten Zusammenfassung wird die Heuristik „Shortcuts“ nicht berücksichtigt. Aufgrund der tendenziell niedrigen ICT-Literacy ist nicht zu erwarten, dass Teilnehmende Shortcuts verwenden, da es Irritationen hervorrufen könnte, wenn aus Versehen eine Tastaturkombination getätigt und ausgeführt wird, und der Effekt von den Teilnehmenden nicht nachvollzogen werden kann.

Die Individualisierbarkeit wird nur hinsichtlich gängiger Anpassungsmöglichkeiten (Größe und Lautstärke) umgesetzt. Weitere Anpassungsmöglichkeiten – wie z. B. die in dem Prinzip „Wahrnehmbar“ der WCAG formulierte Anpassung Kontraste – würden ggf. vom Inhalt ablenken oder die Teilnehmenden überfordern. Vor dem Hintergrund der theoretischen Überlegungen werden folgende Heuristiken für die Zielgruppe der funktionalen Analphabeten und Analphabetinnen als adäquat eingestuft und in weiteren Überlegungen zur empirischen Untersuchung berücksichtigt:

Vorläufige Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen

1. Erlernbarkeit
2. Konsistenz
3. Feedback
4. Selbstbeschreibungsfähigkeit
5. Fehlertoleranz
6. Individualisierbarkeit
7. Wahrnehmungssteuerung
8. Aufgabenangemessenheit
9. Erwartungskonformität
10. Steuerbarkeit
11. Perspektivübernahme
12. Prozessangemessenheit
13. Datensicherheit
14. Joy of Use
15. Interkulturelle Aspekte

Tabelle 6: Vorläufige Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen

Zusammenfassend ist festzuhalten, dass aufgrund der unterschiedlichen Heuristiken und der Abhängigkeit des Anwendungskontextes keine Sammlung oder Systematik von Qualitätskriterien oder Anforderungen an die Usability den Anspruch der Allgemeingültigkeit besitzen kann. Eine adäquate Usability für eine Online-Testumgebung ist gegeben, wenn die Gestaltung des Interfaces weder durch Handhabungs- noch durch Funktionsprobleme die Lösung der Aufgabe beeinflusst. Für die Gewährleistung dieses Anspruchs ist es unabdingbar, ausführliche iterative Evaluationen sowohl mit Usability-Experten und -Expertinnen als auch mit potenziellen Nutzern und Nutzerinnen durchzuführen. Auf der Grundlage von Evaluationen gilt es im empirischen Teil dieser Arbeit u. a., die Heuristiken zu überprüfen und ggf. anzupassen bzw. weitere Heuristiken aufzustellen.

2.7 Evaluation von Usability

Ziel der Usability-Evaluation ist, Schwierigkeiten und Probleme im Umgang mit dem System zu identifizieren. Usability-Evaluation bewertet nicht das System insgesamt – beispielsweise hinsichtlich seiner Relevanz im Bildungswesen – sondern danach, ob bestimmte Kriterien erfüllt sind. Ohne die Evaluation eines Produkts ist nicht festzustellen, inwieweit das Angebot den Bedürfnissen, Erwartungen und Fähigkeiten der potenziellen Zielgruppe entspricht. Zumeist verläuft die Usability-Evaluation – bzw. das Usability-Engineering, welches den Entwicklungs- und Modifikationsprozess bezeichnet – formativ. Das bedeutet, es gibt aufeinanderfolgende Designzyklen, mit der eine Datenbasis geschaffen wird, auf deren Grundlage Iterationen zur Verbesserung der Effektivität, der Effizienz und der Zufriedenheit vorgenommen werden (vgl. zur formativen Evaluation auch Kap. 2.1 und Kap. 3 zum DBR-Ansatz). Einzelne oder mehrere Phasen des Entwicklungsprozesses werden so lange in Zyklen durchlaufen bis die vorher definierten Entwicklungsziele erreicht wurden.

„Usability-Evaluation basiert auf dem Prinzip des nutzer-/nutzerinnenfreundlichen Designs“ (Niegemann, 2008, S. 426). Voraussetzung für die Evaluation ist die Erstellung eines Interface-Entwurfs, die Festlegung des Nutzungskontextes, die Erstellung einer Anforderungsanalyse sowie die Bestimmung der potenziellen Zielgruppe. Es ist festzustellen, was potenzielle Nutzer und Nutzerinnen als gebrauchstauglich und nutzer-/nutzerinnenfreundlich empfinden und wie sie mit dem System beabsichtigen zu arbeiten. Daraus werden des Weiteren notwendige Funktionen abgeleitet und definiert. Die Grundlage für Aussagen und Bewertungen der Usability sind die mittels empirischer Forschungsmethoden erhobenen Daten (Niegemann, 2008, S. 426). Für die Evaluation der Usability bestehen unterschiedliche Methoden, die nach unterschiedlichen Ansätzen systematisiert werden. Die Systematisierungen der Methoden unterscheiden sich einerseits in ihrem Grad der Differenzierung und andererseits hinsichtlich des Unterscheidungskriteriums.

Fisk et al. (2009) wählen einen geringen Differenzierungsgrad auf Grundlage des Kriteriums, ob Nutzer und Nutzerinnen einbezogen werden und differenzieren zwischen User Centered-Methoden („Thinking aloud“, Zielgruppenanalyse, Pilot-Testing, Feldbeobachtung, Fragebögen, Fokus-Groups) und Non User Centered-Methoden (Check-Listen, Heuristische Evaluation, Layout Analysen). Der Ansatz nach Sarodnick und Brau systematisiert anhand des Kriteriums Evaluationsmethoden und differenziert nach formal-analytischen Verfahren, Inspektionsmethoden, Usability-Tests und dem Einsatz von Fragebögen (Sarodnick & Brau, 2011). Der formal-analytische Ansatz wird im folgenden Kap. 2.7.1 näher erläutert. Der „reine“ Usability-Test wird im Kap. 2.7.5 erläutert. Inspektionsmethoden sind wiederum eine generelle Bezeichnung für ein Set an Methoden,

welche den Einsatz von Evaluierenden, die das Interface „inspizieren“, beinhaltet. Zu den Inspektionsmethoden zählen Heuristische Evaluationen und der Cognitive Walkthrough (vgl. auch Kap. 2.7.2). Als weitere Methoden, auf die aber nicht weiter eingegangen wird, werden „formal usability inspections“, „pluralsite walkthroughs“, „feature inspections“, „consistency inspection“ sowie „Standard inspection“ genannt (Sarodnick & Brau, 2011).

Nielsen (1994) zeigt vier grundlegende Wege auf, die Usability zu evaluieren, wobei die Unterscheidungskriterien die Erhebungsmethoden und -instrumente darstellen: automatisch (beispielsweise über das Loggen von Daten), empirisch (über die Testung des Produkts mit realen Nutzern und Nutzerinnen), formal (mit Hilfe von Modellen und formalen Anforderungen, auf deren Grundlage die Usability bewertet wird) sowie informell (basierend auf „Daumenregeln“ und der Fähigkeiten und Erfahrungen der Evaluatoren und Evaluatorinnen bzw. Entwickler und Entwicklerinnen). Nielsen bewertete allerdings zur Zeit dieser Systematisierung die automatischen Methoden als nicht umsetzbar und die formalen Methoden als zu aufwendig – insbesondere bei der Überprüfung von Produkten mit mehreren und/oder komplexen Benutzer-/Benutzerinnenschnittstellen (Nielsen, 1994, S. 413). Inzwischen sind automatisierte Methoden geläufig und werden vielfach eingesetzt.

Auch Niegemann u. a. (2008) differenzieren zwischen vier Kategorien der Evaluation: 1. dem formal-analytischen, 2. dem produktzentrierten, 3. dem interaktionszentrierten Messansatz sowie 4. dem benutzer-/benutzerinnenorientierten Messansatz (Niegemann, 2008, S. 430). Der Ansatz geht auf Bevan u. a. zurück (1991). Das Kriterium ist hierbei der Untersuchungsgegenstand unter Berücksichtigung des Entwicklungsstadiums.

Dieser Differenzierungsansatz spiegelt die beiden grundsätzlichen Untersuchungsgegenstände dieser Arbeit systematisch wieder, indem die Zielgruppe der funktionalen Analphabeten und Analphabetinnen durch die Methode des benutzer-/benutzerinnenorientierten Messansatzes sowie das computerbasierte Diagnoseinstrument durch die anderen drei Messansätze fokussiert werden. Dieser Systematisierungsansatz wird daher im Folgenden genutzt, um die Methoden der Usability-Evaluation darzustellen. Auch dient er als Grundlage des weiteren Vorgehens.

Im Folgenden werden die vier Ansätze erläutert. Der Fokus liegt dabei auf dem produktzentrierten, dem interaktionszentrierten und dem benutzer-/benutzerinnenzentrierten Messansatz, da diese in der Praxis häufig Anwendung finden und auch zur Beantwortung der Forschungsfrage dieser Arbeit herangezogen werden.

2.7.1 *Formal -Analytischer Messansatz*

Der formal-analytische Ansatz wird insbesondere in frühen Phasen der Entwicklung eingesetzt und bewertet die Usability auf der Grundlage formaler Eigenschaften der potenziellen Nutzer und Nutzerinnen (Performanz sowie psychomomentale Leistungen) und des Produktes. Beispielsweise wird (nach dem sog. Keystroke-Level-Modell) im Vorfeld antizipiert, welche Handlungen notwendig sind, um bestimmte Aufgaben zu erfüllen (z. B. Eingabe in die Tastatur) und wie viel Zeit dafür benötigt wird. Demnach kann dazu auch das Antizipieren des Verhaltens von potenziellen Nutzern und Nutzerinnen gehören. Hierfür werden typische Charakteristiken der Zielgruppe identifiziert (beispielsweise über Fragebögen) und in die weitere Entwicklung und Evaluation einbezogen. Diese Angaben werden mit erhobenen Daten in Usability-Tests verglichen. Systematische Abweichungen sollen auf Usability-Schwächen hinweisen. Häufige Kritikpunkte sind die Beschränkung auf den Zeitfaktor sowie der fehlende Einbezug von den Usability-Attributen Effektivität und Zufriedenheit (Niegemann, 2008, S. 429).

2.7.2 *Produktzentrierter Messansatz*

Beim produktzentrierten Messansatz steht das Produkt im Vordergrund. Es werden ergonomische Eigenschaften der Anwendung gemessen und beurteilt. Diese können wiederum differenziert werden in Inspektionsmethoden sowie Fragebögen und Checklisten (vgl. z. B. Niegemann, 2008, S. 200; Nielsen & Loranger, 2008).

Shneiderman und Plaisant (2005) fassen produktzentrierte Messansätze unter dem Begriff Expert-Reviews zusammen. Expert-Reviews sind Begutachtungen der Umgebung durch Experten und Expertinnen. Hierzu zählen a) die Heuristische Evaluation, b) das Guidelines Review, c) die Consistency Inspection, d) der Cognitive Walkthrough sowie e) die Formal Usability Inspection. Die Methode der Heuristischen Evaluation wird im Folgenden ausführlicher im Vergleich zu den weiteren Methoden dargestellt, um im empirischen Teil dieser Arbeit auf Teilaspekte der Heuristischen Evaluation zurückgreifen zu können.

a) Heuristische Evaluation

Bei der Methode der Heuristischen Evaluation handelt es sich um Discount-Usability-Engineering mit qualitativem Charakter (Sarodnick & Brau, 2011, S. 144). Bei der Durchführung werden Usability-Prinzipien berücksichtigen, die sich beispielsweise auf die Einhaltung von Konsistenz beziehen.

Bereits 1985 formulierten Gould und Lewis die drei Prinzipien eines guten Designs für nützliche und leicht handhabbare Computersysteme:

„Early Focus on User Tasks, Empirical Measurement und Iterative Design“ (Gould, 1985, S. 300).

Diese drei Prinzipien finden sich in jeglichen Heuristischen Evaluationen wieder. Eine Heuristische Evaluation gehört zu den informellen analytischen Verfahren und Inspektionsmethoden. Hierbei beurteilen eine geringe Anzahl von Experten und Expertinnen anhand von Heuristiken die Usability des zu analysierenden Produktes (Niegemann, 2008, S. 434). Der Erfüllung typischer Guidelines folgen vielzählige zu befolgende Regeln und rufen bei Entwicklern und Entwicklerinnen eher einen einschüchternden Effekt hervor als dass sie ambitioniert befolgt werden. Daher wählen Entwickler und Entwicklerinnen oft den Weg, Heuristische Evaluationen auf der Grundlage ihrer eigenen für sie relevanten Heuristiken durchzuführen (Nielsen & Molich, 1990, S. 249). Eine Heuristische Evaluation wird durchgeführt, indem Experten und Expertinnen ein Interface beobachten und dieses bewerten. Die Bewertung wird meist sinnvollerweise anhand von Kriterien – Heuristiken – vorgenommen. Einerseits kann durch die Nutzung von Kriterien aus der Perspektive der Entwickler und Entwicklerinnen der sie interessierende Fokus gesetzt werden. Andererseits kann dadurch der Vergleich von Beobachtungen verschiedener Personen einfacher gewährleistet werden (Nielsen & Molich, 1990, S. 249). Viel zitierte und angewandte Heuristiken sind die nach Nielsen und Molich (1990) und die „8 golden rules“ nach Shneiderman und Plaisant (2005) (vgl. Kap. 2.6.4.1).

Einen möglichen Ablauf einer heuristischen Evaluation beschreibt Niegemann (2008): Anhand der ausgewählten und ggf. modifizierten oder erweiterten Heuristiken beurteilen die einzelnen Experten und Expertinnen separat die Usability der Anwendung, idealerweise in zwei Durchgängen. Zudem können die Heuristiken im Team bewertet werden. Vor der heuristischen Evaluation sind die Experten und Expertinnen mit den heuristischen Prinzipien vertraut zu machen. Finden zwei Durchläufe statt, konzentrieren sich die Experten und Expertinnen im ersten Durchlauf auf den Informationsablauf und die Funktionalitäten, im zweiten Durchgang auf die einzelnen Bedienelemente. Die Usability-Probleme werden dabei schriftlich festgehalten, indem genau beschrieben wird, welche der heuristischen Prinzipien wie verletzt wurden. Nachdem die Probleme vorerst in Einzelarbeit analysiert und anschließend in der Gruppe der Experten und Expertinnen diskutiert wurden, wird eine Gesamtliste der identifizierten Usability-Probleme erstellt. In dem darauf folgenden Schritt werden die Usability-Probleme nach ihrem Schweregrad durch die Experten und Expertinnen beurteilt. Problematisch ist hierbei, dass es für den Schweregrad keine einheitliche Definition gibt. Folgende Faktoren sollten in die Beurteilung der Usability-Probleme einfließen (Nielsen, 2014): 1. Auftretenshäufigkeiten, 2. Einfluss des

Problems auf die Erreichung des gesetzten Ziels bzw. der Aufgabenbewältigung, 3. Persistenz des Problems.

Sarodnick und Brau (2011) beschreiben eine weitere Art der heuristischen Evaluation: die Kooperative Heuristische Evaluation. Dabei erarbeitet ein unabhängiger Versuchsleiter/eine unabhängige Versuchsleiterin zusammen mit den Entwicklern und Entwicklerinnen realistische Anwendungsszenarien. Usability-Experten und -Expertinnen werden mit dem System vertraut und mit den Soll-Prozessen bekannt gemacht. Während der Evaluation bearbeitet ein Usability-Experte/eine Usability-Expertin die Anwendungsszenarien; die evaluierende Person soll dabei die Schritte kommentieren. Der Usability-Experte/Die Usability-Expertin stellt Verständnisfragen zu Handlungsabfolgen und zu realen Arbeitstätigkeiten. Die Beschreibung realer Arbeitsschritte soll dem Usability-Experten/der Usability-Expertin dabei helfen, die Sicht der Nutzer und Nutzerinnen einzunehmen. Bei der Durchführung einer heuristischen Evaluation sind nach Sarodnick und Brau (2011) drei Faktoren zu berücksichtigen:

1. Der Usability-Experte/Die Usability-Expertin ist gleichzeitig Evaluator/Evaluatorin, Interviewer/Interviewerin und Lernender/Lernende. Die Fähigkeiten, sich während der Gesprächsführung in die Prozesse hineinzuversetzen und entsprechend zu reagieren, haben einen starken Einfluss auf den Evaluationserfolg.
2. Die potenziellen Nutzer und Nutzerinnen müssen die Inhalte der Prozesse verständlich und strukturiert wiedergeben.
3. Pro Räumlichkeit sollte nur eine Evaluation zurzeit stattfinden, da die Evaluierendenpaare sich gegenseitig ablenken oder beeinflussen könnten.

Die Heuristische Evaluation kann unterschiedlich stark detailliert oder umfangreich stattfinden. Es ist möglich, eine Heuristische Evaluation mit sehr wenig Aufwand durchzuführen, je nachdem, wie viele Heuristiken anhand wie vieler Aufgaben überprüft werden sollen. Daher handelt es sich auch um eine Methode des Discount-Engineerings.

b) Guidelines Review

Hierbei werden von Experten und Expertinnen relevante Guidelines erweitert und/oder modifiziert, um auf der Grundlage dieser das Interface hinsichtlich der Konformität mit den relevanten Guidelines zu überprüfen. Im Vergleich zur heuristischen Evaluation wird hier weniger systematisch vorgegangen.

c) Consistency Inspection

Bei dieser Methode wird der Fokus auf die Konsistenz der Interfaces gelegt, insbesondere wird darauf geachtet, dass Symbole, Terminologie, Schriftart, Farbschemata, das Layout sowie Input- und Output-Formate konsistent dargestellt sind.

d) Cognitive Walkthrough

Bei einem Cognitive Walkthrough werden Aufgaben und Handlungen von Experten und Expertinnen simuliert, die von potenziellen Nutzern und Nutzerinnen durchgeführt werden. Hierzu zählt zudem, einen typischen Tagesablauf eines potenziellen Nutzers/einer potenziellen Nutzerin zu antizipieren, um die mögliche Nutzung der Webseiten/des Programms im Kontext der Lebenswelt des Nutzers/der Nutzerin betrachten zu können. Die Methode des Cognitive Walkthrough wurde entwickelt, um insbesondere Interfaces zu evaluieren, deren Umgang explorativ erlernt wird. Cognitive Walkthroughs können durch Einzelne durchgeführt werden, wobei es jedoch sinnvoll ist, Cognitive Walkthroughs von möglichst unterschiedlichen Experten und Expertinnen (Programmierern/Programmiererinnen, Designern/Designerinnen, Nutzer/Nutzerinnen) durchführen zu lassen.

e) Formal Usability Inspection

Experten und Expertinnen präsentieren weiteren Entwicklern und Entwicklerinnen das Interface, um Stärken und Schwächen zu diskutieren und ggf. Lösungen für Probleme zu finden. Diese Form der Evaluation ist personell aufwändig und bleibt auf einer theoretischen Ebene. Ferner gibt sie Personen mit wenig Erfahrung die Möglichkeit, ihr Wissen über Usability-Anforderungen zu erweitern.

Zusammenfassend ist der produktzentrierte Messansatz in jeglichen Entwicklungsstadien anwendbar. Dadurch, dass potenzielle Nutzer und Nutzerinnen für die Durchführung nicht zwingend notwendig sind, kann der produktzentrierte Messansatz mit wenig Aufwand und schnell umgesetzt werden. Der Nachteil liegt jedoch auch genau darin: ohne potenzielle Nutzer und Nutzerinnen ist das Risiko groß, relevante Bedürfnisse von Nutzern und Nutzerinnen nicht zu erkennen und somit nicht berücksichtigen zu können.

2.7.3 Interaktionszentrierter Messansatz

Der interaktionszentrierte Messansatz fokussiert die Interaktion zwischen Mensch und Computer, wobei die Messung der Performanz sowie die psychomentalen Leistungen im Fokus stehen. Hierfür werden Probanden und Probandinnen aufgefordert, mit dem System zu arbeiten und ggf. bestimmte Aufgaben zu bearbeiten. Das Szenario gleicht damit dem eines Usability-Tests (das sogenannte „reine“ Usability-Testing wird in Kap. 2.7.5 erläutert). Es können Beobachtungen, Screen-Recording (die Handlungen des Nutzers/der Nutzerin werden videographiert), Logfile Recording (z. B. Speicherung der Anzahl und Reihenfolge der Seitenaufrufe, Dauer der Nutzung) und/oder Eye-Tracking (Aufnahme der Augenbewegungen) durchgeführt sowie Beobachtungsprotokolle erstellt werden. Eine spezielle Form, die im frühen Entwicklungsstadium eines zu entwickelnden Systems angewendet werden kann, ist das Paper Prototyping.

Paper Prototyping ist eine Methode, um Benutzer-/Benutzerinnenoberflächen zu entwickeln, zu testen und zu verbessern. Snyder (2007) definiert Paper Prototyping als eine Form eines Usability-Tests, in dem repräsentative Nutzer und Nutzerinnen realistische Aufgaben mit einer Papierversion der Benutzer-/Benutzerinnenschnittstelle durchführen (Snyder, 2007, S. 4). Eine Person simuliert dabei das Verhalten des Produktes (z. B. legt die entsprechenden Seiten vor, die erscheinen wenn eine bestimmte Schaltfläche aktiviert wurde) und reagiert nur auf die Aktionen der Nutzer und Nutzerinnen. Die computersimulierende Person liefert keine weiteren Erläuterungen und kommuniziert nicht mit dem Nutzer/der Nutzerin. Im Vorfeld ist ein Aufgabensample auf Papierbasis vorzubereiten. Des Weiteren sind alle Optionen, die durch das Aktivieren von Schaltflächen folgen können, vorzubereiten. Während des Paper Prototypings bekommt der Nutzer/die Nutzerin eine Papiermaus in die Hand, mit der er/sie eine Computermouse simuliert. Simuliert sie das Anklicken einer Schaltfläche hat die computersimulierende Person das Interface vorzulegen, welches erscheinen würde, wenn das Produkt entwickelt und programmiert ist.

Zur Teilnahme am Paper Prototyping sind potenzielle Nutzer und Nutzerinnen auszuwählen, die über das Ziel des Experiments, die Rahmenbedingungen und die Spielregeln informiert werden. Zudem werden sie aufgefordert, während des Experiments laut zu denken, um Aufschluss über weitere Aspekte zu erhalten, die womöglich nicht bedacht wurden. Vor der Durchführung sind die Rollen zu definieren und zu verteilen (Computer simulierende Person, Testperson bzw. Nutzer/Nutzerin, Beobachter/Beobachterin, Unterstützer/Unterstützerin für die Bereitstellung der Materialien). Die Dokumentation erfolgt idealerweise neben der Beobachtung über eine Videoaufzeichnung. Im Anschluss an das Experiment kann ein Interview mit dem Proband/der Probandin durchgeführt werden, um Informationen über die Wahrnehmung, Erfahrungen und Emotionen zu erhalten (Wolf & Koppel, 2010, S. 225). Die Methode dient zur Erstellung von Interface-Entwürfen durch interdisziplinäre Design-Teams (Hornecker, 2004) ebenso wie zur schnellen Evaluation von Entwürfen (Hackos, 1998; Rogers u. a., 2011).

Der Fokus liegt zwar auf dem Verhalten des Nutzers/der Nutzerin bzw. des/der Lernenden, es werden jedoch weniger deren Urteile erfragt (wie bei den benutzer-/benutzerinnenorientierten Messansätzen – vgl. Kap. 2.7.4), sondern ihre Handlungen protokolliert und später analysiert. Stehen Wahrnehmung und das Erleben der Nutzer und Nutzerinnen im Mittelpunkt des Interesses, werden benutzer-/benutzerinnenorientierte Messmethoden eingesetzt.

2.7.4 Benutzer-/Benutzerinnenorientierte Messansatz

Der benutzer-/benutzerinnenorientierte Messansatz wird angewendet, wenn primär die subjektive Beurteilungen potenzieller Nutzer und Nutzerinnen erhoben

werden sollen. Methoden zu dessen Feststellung sind Thinking Aloud (Lewis, 1982), Videokonfrontationsmethode (Neal & Simons, 1984), Fokus Groups (Nielsen, 1997), Fragebögen und Checklisten, sowie Usability-Tests mit dem Fokus auf die Wahrnehmung der Nutzer und Nutzerinnen. Für die Wahrnehmung der Nutzer und Nutzerinnen – die UE (vgl. in Kap. 2.6.2) – existieren standardisierte Instrumente; z. B. der AttrakDiff (Hassenzahl, Burmester & Koller, 2003) und der User Experience Questionnaire (Laugwitz, Schrepp & Held, 2006). Eine weitere Methode ist die Durchführung von Interviews – bei Niegemann auch als Question Asking bezeichnet (Niegemann 2008, S. 242). Diese Methode wird näher erläutert, da diese im empirischen Teil dieser Arbeit Anwendung fand. Eine umfassende Erläuterung wird in Kap. 8.2.2.1 vorgenommen.

Interviews können dazu dienen, weitere Aspekte aus der Perspektive der Nutzer und Nutzerinnen zu erfragen und Aufschluss über deren Wahrnehmung, Empfindungen und Einschätzung des Produktes zu erhalten. Da die Interviewmethode primär offene Frageformate verwendet, bietet sie z. B. auch die Möglichkeit, Verbesserungsvorschläge aufzunehmen. Im Vorfeld sollte dabei entschieden werden, welche Form von Interview durchgeführt werden sollte. Je nach Zielgruppe können unterschiedliche Methoden sinnvoll sein. So kann bei einem leitfadengestützten Interview generell die Wahrnehmung oder das Nutzungsverhalten von Personen im Vordergrund stehen. Bei fokussierten Interviews steht ein Gegenstand im Fokus, der beiden Interviewteilnehmenden bekannt ist. Da bei Usability-Tests meistens Informationen über die Wahrnehmung eines bestimmten Produktes und ggf. bestimmter Aspekte des Produktes erfragt werden sollen, eignet sich insbesondere die Methode des fokussierten Interviews. Hierfür ist im Vorfeld ein Leitfaden zu entwickeln und zu erproben. Zudem ist sicherzustellen, dass die Interviewer und Interviewerinnen die Fragetechniken beherrschen, um die interviewte Person nicht zu beeinflussen, z. B. durch Suggestivfragen (vgl. z. B. Helfferich 2004 und Kap. 8.2.2.1).

2.7.5 Usability-Tests in Laboratories

Usability-Tests werden durchgeführt, wenn das System in einem Entwicklungsstadium existiert, welches die Bearbeitung von realen Aufgaben durch Nutzer und Nutzerinnen ermöglicht. Mit Usability-Tests können Daten akquiriert werden, die sowohl Aufschluss über die Interaktion als auch über die potenziellen Nutzer und Nutzerinnen sowie die Passung von Mensch und System liefern. Damit sind auch die Fähigkeiten des potenziellen Nutzers/der potenziellen Nutzerin impliziert (vgl. z. B. Bevan u. a., 1991). Je nach Fokus können sie daher sowohl dem interaktions- als auch dem benutzer-/benutzerinnenorientierten Messansatz zugeordnet werden. Liegt der Fokus auf der Performanz und dem

System bzw. der Interaktion mit dem System, ist der Usability-Test dem interaktionszentrierten Messansatz zuzuordnen. Wird die Wahrnehmung und das Erleben mit dem System fokussiert, wird diese Methode dem benutzer-/benutzerinnenorientierten Messansatz zugeordnet (Niegemann, 2008, S. 444). Bei Usability-Tests in Laboratories werden Evaluationen in kontrollierter Umgebung – in Laboren – durchgeführt, um statistisch signifikante Unterschiede bezüglich der optimalen Gestaltung der Benutzer-/Benutzerinnenoberfläche zu erhalten. Dies geschieht dadurch, dass mindestens zwei experimentelle Bedingungen geprüft und verglichen werden. Hierfür werden Aufgaben ausgewählt, die hinsichtlich bestimmter Bedingungen variieren und von potenziellen Nutzern und Nutzerinnen durchgeführt werden. Damit soll festgestellt werden, welche Änderungen vorzunehmen oder – sollten Hypothesen vorab formuliert worden sein – welche Hypothesen zu verifizieren bzw. zu falsifizieren sind.

Bezüglich der Erhebung der subjektiven Beurteilung ist zu berücksichtigen, dass Maskierungseffekte auftreten können (Bevan u. a., 1991). So sind Usability-Tests in kontrollierter Umgebung keine Abbildung der Realität und geben somit ggf. nur bedingt Aufschluss darüber, wie sich die Nutzer und Nutzerinnen tatsächlich verhalten würden.

Für eine möglichst umfassende Problemanalyse und -aufdeckung sind weniger Kenntnisse der technischen Hintergründe entscheidend, sondern vielmehr domänenspezifische Kenntnisse. Nach Nielsen (1992) ist die Problemanalyse effektiver, wenn die Evaluierenden sowohl Domänen- als auch Usability-Expertise aufweisen. Nielsen spricht ihnen eine durchschnittliche Fehlerrückmeldung von 60% zu. Wird die Evaluation von Personen durchgeführt, die jeweils nur in einem der Bereiche Kenntnisse vorweisen, werden lediglich 22% bis 41% der Fehler entdeckt (1992, S. 376).

2.7.6 *Planung und Durchführung von Usability-Evaluation*

Bereits 1985 formulierte Shackel fünf Grundsätze für den Designprozess der Usability: „User-Centered Design“ (Designer und Designerinnen müssen Kenntnisse über die potenzielle Nutzer-/Nutzerinnengruppe und deren Handlungsinteresse haben), „Participative Design“ (ein Gruppe von potenziellen Nutzern und Nutzerinnen sollte eng mit dem Design-Team zusammenarbeiten und bereits früh in den Entwicklungsprozess einbezogen werden), „Experimental Design“ (potenzielle Nutzer und Nutzerinnen testen Papierprototypen und Prototypen in frühen Entwicklungsstadien), „iterative Design“ (Schwierigkeiten, die in früheren Entwicklungsstadien identifiziert wurden müssen modifiziert und einer weiteren Prüfung unterzogen werden) und „User supportive design“ (Unterstützende Funktionen für Nutzer und Nutzerinnen sind in frühe Entwicklungsprozesse einzubeziehen und in frühen Experimenten zu überprüfen) (Shackel, 1985, S. 20).

ff). Um aussagekräftige Ergebnisse zu erhalten, ist neben den zu berücksichtigenden Grundsätzen eine detaillierte Planung der Test-Durchführung notwendig, die folgende Schritte beinhaltet (angelehnt an Niegemann, 2008, S. 446 ff):

- 1) *Detaillierte Planung*: Definition von Testzweck, Testgegenstand, Zielgruppe, Auswahl von Erhebungsmethode und -technik, Räumlichkeiten und Testinhalt – z. B. das Aufgabensample – sowie die Entscheidung darüber, welche Daten erhoben und wie sie ausgewertet werden sollen. Auch ist eine Zeit- und Kostenabschätzung vorzunehmen.
- 2) *Zusammenstellung des Testmaterials und der Testumgebung*: Neben dem Untersuchungsgegenstand sind weitere Testmaterialien vorzubereiten. Dies können z. B. Fragebögen, Interviewleitfäden und psychologische Tests sein, die Aufschluss über die Wahrnehmung, das Empfinden, Gefallen und die Kompetenz der Nutzer und Nutzerinnen geben. Nach der Auswahl der Aufgaben sind die Szenarien vorzubereiten, es sind Briefingmaterial, Ablaufübersicht und weitere relevante Informationen für die Teilnehmenden zusammenzustellen. Auch sollten die Instruktionen für die Testleiter schriftlich vorliegen. Wichtig für die Verwendung der Daten ist die zu unterschriebene Einverständniserklärung der Teilnehmenden.
- 3) *Durchführung der Usability-Tests*: Je nach ausgewählter Methoden bzw. Kombination von Methoden wird der Usability-Test durchgeführt. Jede teilnehmende Person bearbeitet hierfür ein im Vorfeld definiertes Aufgabensample. Hierfür können zusätzlich Aufzeichnungsmethoden wie Videoaufnahme, Screen-Recording, Mouse-tracking und/oder Eye-Tracking verwendet werden.
- 4) *Instruktion der Teilnehmer und Teilnehmerinnen*: Die Teilnehmenden müssen über den Ablauf, das Ziel des Usability-Tests und Umgangsmöglichkeiten bei Problemen instruiert werden. Auch ist ihnen die Möglichkeit einzuräumen, Fragen zu stellen und – falls vorgesehen – spezielle Methoden wie beispielsweise lautes Denken einzuüben.
- 5) *Datensammlung und Analyse*: Nach der Datenerhebung müssen die Daten systematisch aufbereitet werden. Protokolle und Beobachtungen müssen schriftlich ausgearbeitet und für die Auswertung vorbereitet werden.
- 6) *Berichtlegung und Zusammenstellung*: Im letzten Schritt werden Empfehlungen zur Modifikation des Produktes in einem Ergebnisbericht zusammengestellt. Dabei sollte die Darstellung der Modifikationsnotwendigkeiten im Vordergrund stehen. Zu dem Bericht zählen die Vorstellung und Erläuterung der angewendeten Methoden und Techniken so-

wie die Darstellung der Ergebnisse und eine Auflistung der Empfehlungen.

Im empirischen Teil dieser Arbeit sind die eben dargestellten Grundsätze sowie die Ablaufschritte dahingehend umzusetzen, dass im Sinne des Grundsatzes „User-Centered Design“ zielgruppenspezifische Charakteristika berücksichtigt (vgl. Kap. 2.3 zu zielgruppenspezifischen Voraussetzungen im Theorieteil) sowie zielgruppenunterstützende Funktionen entwickelt und überprüft werden. Entsprechend des „Participative Design“ sind potenzielle Nutzer und Nutzerinnen bereits in frühen Entwicklungsstadien einzubeziehen, indem von ihnen erste Prototypen getestet werden („Experimental Design“). Des Weiteren ist bei dem Entwicklungsprozess der computerbasierten Diagnostik für funktionale Analphabeten und Analphabetinnen iterativ („Iterative Design“) in mehreren Entwicklungszyklen vorzugehen, so dass der Ablauf einer Usability-Evaluation – wie eben in Anlehnung an Niegemann dargestellt – nicht „singulär-linear“ erfolgt, sondern in mehreren Zyklen, in denen die einzelnen Schritte wiederholt werden. Den Rahmen für das empirische Vorgehen bildet der DBR-Ansatz (vgl. Kap. 3). Die Grundsätze der Usability-Evaluation werden in diesen Rahmen systematisch integriert und miteinander verknüpft.

2.8 Zusammenfassung der theoretischen Herleitung

Ausgehend von dem Alphabetisierungsbedarf wurden die aktuelle Situation funktionaler Analphabeten und Analphabetinnen sowie die Herausforderungen in der Alphabetisierung erläutert. Zur Reduzierung des funktionalen Alphabetismus kann eine computerbasierte Diagnostik beitragen, die eine effiziente Kompetenzdiagnostik und die Ableitung individueller Fördermaßnahmen ermöglicht. Die technische Entwicklung bietet Möglichkeiten der automatisierten Auswertung von offenen Antwortformaten, wodurch z. B. über die Auswertung von Textpassagen die Schreibkompetenzen von funktionalen Analphabeten und Analphabetinnen diagnostiziert werden können. Ziel des Forschungsvorhabens ist die Entwicklung und Evaluation einer Online-Testumgebung für funktionale Analphabeten und Analphabetinnen mit Single Choice- und (halb-)offenen Antwortformaten. Bei der Zielgruppe der funktionalen Analphabeten und Analphabetinnen ist allerdings von einer niedrigen ICT-Literacy auszugehen. Folglich ist das computerbasierte Diagnoseinstrument entsprechend gebrauchstauglich zu gestalten. Hinweise für die Gestaltung der Testumgebung liefern die Erkenntnisse zur Usability sowie zur CLT und CTML. Die folgende Abbildung stellt die Wirkungszusammenhänge der für die Forschungsfrage relevanten Faktoren reduziert dar:

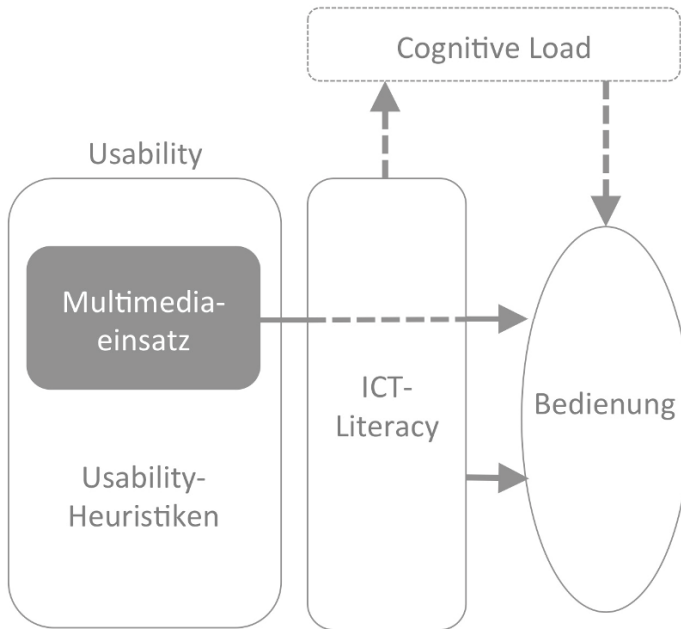


Abbildung 7: Wirkungsbeziehungen basierend auf der theoretischen Herleitung

Die durchgezogenen Linien verdeutlichen den vermuteten direkten Zusammenhang, die perforierten Pfeile weisen auf einen indirekten bzw. indirekt beobachtbaren Zusammenhang hin. Entsprechend der dargestellten theoretischen Basis ist die Usability eines Produktes durch die Effektivität, die Effizienz und Zufriedenheit charakterisiert. Zur Erreichung dieser Attribute dienen Usability-Guidelines und –Heuristiken. Der Einsatz von Multimedia kann zusätzlich zu einer ausgeprägten Usability beitragen. Im Kontext dieses Forschungsvorhabens wird die Usability über die Usability-Heuristiken und zudem den Einsatz von multimedialen Unterstützungsfunktionen evaluiert. Über das Bedienverhalten potenzieller Nutzer und Nutzerinnen können Rückschlüsse auf die Usability gezogen werden. Zu berücksichtigen ist dabei die tendenziell niedrige ICT-Literacy der funktionalen Analphabetinnen: Es ist davon auszugehen, dass die ICT-Literacy das Bedienverhalten moderiert und einen Einfluss auf die Beanspruchung des Cognitive Load – insbesondere des *Extraneous Load* – hat. Je nach Einsatz von Multimedia und der Ausprägung der Usability ist zu erwarten, dass die ICT-Literacy die Effekte, die schließlich über die Bedienung beobachtbar werden, beeinflusst: Beispielsweise ist bei einer niedrigen ICT-Literacy zu

vermuten, dass eine für den Nutzungszweck mangelnde Gestaltung einen stärkeren Einfluss auf das Bedienverhalten hat als Personen mit einer hohen ICT-Literacy. Als Zusatzexplanation dienen Erkenntnisse zur CLT: Beispielsweise ist bei einer niedrigen ICT-Literacy davon auszugehen, dass der *Extraneous Load* durch eine nicht nutzer-/nutzerinnenfreundliche Gestaltung stärker beansprucht ist als bei Personen mit einer hohen ICT-Literacy. Der Einsatz von Multimedia kann demnach die Reduktion der Beanspruchung des Cognitive Load unterstützen und die Bedienung der Online-Testumgebung beeinflussen.

Gemäß des DBR-Ansatzes (vgl. Kap. 3) wurde mit dieser theoretischen Herleitung der Alphabetisierungsbedarf und die mangelnden Ressourcen zur Reduzierung des Bedarfs als Problem in der Praxis identifiziert. Mit Hilfe aktueller Forschungserkenntnisse wurde das Forschungsvorhaben konkretisiert und hinsichtlich seiner praktischen sowie theoretischen Relevanz dargestellt. Damit ist die erste Phase des DBR-Prozesses abgeschlossen. Vor dem Hintergrund der bisher entfalteten theoretischen Basis (Computerbasierte Diagnostik, zielgruppenspezifische Voraussetzungen, CLT sowie der CTML und Usability) wurde die Online-Testumgebung *otu.lea* konzipiert und mit dem Design-Based Research-Ansatz evaluiert. In der folgenden Phase (Phase II) werden sowohl die Konzeption der Online-Testumgebung als auch die Konzeption des Forschungsrahmens erläutert. In der darauf folgenden Phase (Phase III) wird die Evaluation der Online-Testumgebung vorgestellt und es werden die in diesem Kapitel dargestellten Wirkungszusammenhänge im Kontext des Einsatzes der Online-Testumgebung für funktionale Analphabeten und Analphabetinnen untersucht. Abschließend erfolgt eine Reflexion des Prozesses (Phase IV).

PHASE II – Konzeption

3 Konzeption des Forschungsrahmens - Design-Based Research (DBR)

*„Educational researchers are increasingly using design as a means of advancing their understanding“
(Edelson, 2002, S. 105).*

Design-Based Research (DBR) wird im erziehungswissenschaftlichen Kontext zunehmend angewendet (T. Anderson & Shattuck, 2012, S. 19) und für die Untersuchung und Entwicklung von neuen Lehr- und Lerntechnologien als gewinnbringend, im Vergleich zu klassischen Forschungsansätzen, bezeichnet (Reeves, 2006). Insbesondere in den USA findet dieser Ansatz Zuspruch (T. Anderson & Shattuck, 2012, S. 20). In der Literatur werden verschiedene Begrifflichkeiten für Design-Based-Research (DBR) verwendet: design experiments (A. L. Brown, 1992; Collins, 1992), design research (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003; Collins, Joseph & Bielaczyc, 2004; Edelson, 2002; Middleton, Gorard, Taylor & Bannan-Ritland, 2008) sowie development research (Richey & Klein, 2005). Auch wenn die Begrifflichkeiten meist synonym verwendet werden, gibt es Bemühungen, die Differenzen zwischen diesen Bezeichnungen herauszuarbeiten (Wang & Hannafin, 2005, S. 7). Design experiments haben demnach zum Ziel, verschiedene Aspekte bzw. Konzepte miteinander zu vergleichen (Collins, 1992). *Design Research* hat insbesondere den Output und Outcome im Fokus (Edelson, 2002). Dieser wird zwar auch im DBR-Ansatz verfolgt, doch steht hingegen im Kontext des DBR insbesondere der iterative Charakter im Vordergrund, dessen Prozess aus einzelnen Design-Zyklen besteht. Da das Durchlaufen einzelner Design-Zyklen ein wesentliches Merkmal dieser Forschungsarbeit aber auch ein grundsätzliches Merkmal in der Usability-Entwicklung darstellt, wird im Folgenden der Begriff DBR verwendet.

„[design-based research is] a systematic but flexible Methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real-world settings, and leading to contextually sensitive design principles and theories“ (Wang & Hannafin, 2005, S. 6).

DBR bedient sich unterschiedlicher wissenschaftlicher Methoden, die flexibel miteinander kombinierbar sind und stellt aufgrund der Flexibilität eher ein Forschungsrahmen als eine Evaluationsmethode dar (Wang & Hannafin, 2005, S. 9). Der DBR-Ansatz geht auf das Jahr 1992 zurück und ist damit ein noch recht junger Forschungsansatz. Damals wurden der Design-Begriff und die Idee des „Design-Experiments“ von Ann Brown in der Lehr-Lernforschung eingeführt (A. L. Brown, 1992). Ursächlich war die Suche nach Möglichkeiten, die Lehr-Lernforschung zum Zwecke der Innovation zu betreiben. Daraus entwickelte sich das Bedürfnis nach einem Forschungsansatz, der Lernphänomene in realen Situationen untersucht, dabei über enge Messkriterien hinausgeht und das Design in den wissenschaftlichen Prozess aufnimmt (Reinmann, 2005, S. 60). Zugrunde liegt hierbei mehr ein modernes als ein traditionelles Innovationsverständnis. Im Gegensatz zum traditionellen Innovationsverständnis, welches von dramatischen Effekten und radikalen Veränderungen ausgeht, sind dem modernen Innovationsverständnis inkremental-evolutionäre Neuerungen und eher undramatische Effekte implizit (Reinmann, 2005, S. 53).

Im erziehungswissenschaftlichen Kontext ist die Motivation, den Design-Based-Research-Ansatz anzuwenden, meist in dem Streben nach kontinuierlicher Verbesserung der Bildungspraxis begründet. Hierbei spielen die enge Verbindung zwischen Theorieentwicklung, Optimierung von Gestaltungsprozessen und insbesondere der enge Bezug zur Praxis eine große Rolle (Cobb u. a., 2003).

Vorwiegend wird der DBR-Ansatz in Lehr-Lernkontexten angewandt. Wenn neue Technologien und das Potenzial dieser im Lehr-Lernkontext erforscht werden, haben das Vorgehen und die Einsicht in den Prozess mitunter Veränderungen des Denkens über den Lehr-Lernkontext bzw. -inhalt zur Folge. Der Ansatz ermöglicht die Erweiterung oder die Neuformulierung von Theorien im pädagogischen Kontext. Dies kann ebenso durch den Einsatz neuer Lernmaterialien und Technologien als auch komplexerer Lerneinheiten – beispielsweise in Form von Seminaren und Methoden – realisiert werden.

Im Folgenden werden die theoretischen Grundlagen des DBR-Ansatzes erläutert, indem die Ziele, die Merkmale, der Ablauf sowie das forschungsstrategische und -methodische Vorgehen dargestellt werden. Darauf folgt eine Abgrenzung zu weiteren empirischen Ansätzen, um schließlich den DBR-Ansatz im Anwendungskontext vorzustellen und einer kritischen Reflexion zu unterziehen. In einem späteren Kapitel (Kap. 5.1) erfolgt der Transfer des DBR-Ansatzes auf das Forschungsvorhaben dieser Arbeit.

3.1 Ziele des DBR-Ansatzes

Der DBR-Ansatz verfolgt das Ziel, Lösungen für Probleme in der Bildungspraxis zu finden und nachhaltige Innovationen zu entwickeln. Mit dem DBR-Ansatz

soll dem Problem begegnet werden, dass Lehr-Lernsituationen wesentlich komplexer sind, als dass sie durch die Herstellung eines für Experimente tauglichen Variablen-Sets nachgebildet werden könnten:

*„Ziel ist, durch systematische Gestaltung, Durchführung, Überprüfung und Re-Design genau diese Komplexität besser als bisher zu durchdringen“
(Reinmann, 2005, S. 61).*

Das übergeordnete Ziel ist, kommunizierbare Theorien zu entwickeln, die kontextsensitiv sind und gleichzeitig wissenschaftliche Erkenntnisse liefern (Reinmann, 2005, S. 62).

Um diese Ziele zu erreichen erfolgt auf der theoretischen Grundlage die Entwicklung eines Designs, welches in der Praxis eingesetzt, evaluiert und modifiziert wird (Edelson, 2002; Wang & Hannafin, 2005). Das Design besitzt dabei einen hohen Stellenwert, da der Designprozess selbst Gegenstand des Forschungsprozesses und die Quelle für die Entwicklung von Theorien ist (Reinmann, 2005). Hierbei nehmen bereits im Vorfeld die Kontextfaktoren Einfluss auf die Gestaltung und sind somit von vornherein im Entwicklungs- und Forschungsprozess eingebettet (s. Abbildung 8).

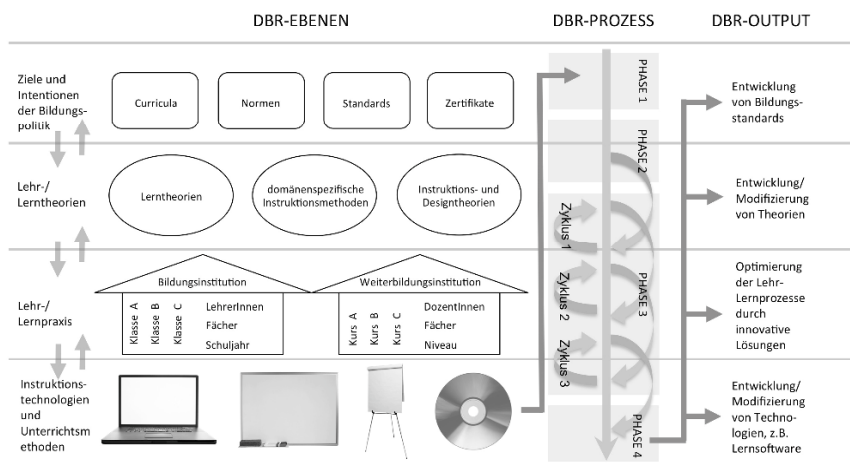


Abbildung 8: DBR-Ebenen, -Prozess und -Ablauf

Im bildungswissenschaftlichen Kontext sind im DBR-Prozess im Wesentlichen vier Ebenen involviert, die tendenziell hierarchisch absteigend zu betrachten sind

und sowohl direkt als auch indirekt aufeinander einwirken: a) politische Ebene: Ziele und Intentionen der Bildungspolitik, b) theoretische Ebene: Lehr-/Lerntheorien, c) praktische Ebene: Lehr-/Lernpraxis und d) methodische Ebene: Instruktionstechnologien und Unterrichtsmethoden.

- a) und b) Die Ziele und Intentionen der Bildungspolitik schlagen sich u. a. in Curricula, Normen, Standards und Zertifikaten nieder. Diese werden vor dem Hintergrund aktueller Theorien und Erkenntnisse formuliert.
- b) und c) Unter Berücksichtigung der Lehr- und Lerntheorien werden schließlich die Ziele und Intentionen der Bildungspolitik in der Praxis umgesetzt. Die Erfahrungen aus der Bildungspraxis schlagen sich wiederum in der (Weiter)Entwicklung von Lehr- und Lerntheorien nieder. Angemerkt sei hier, dass die Ebene der Lehr-/Lerntheorien und der Lehr-/Lernpraxis weniger hierarchisch und eher auf einer Ebene anzusiedeln sind, da diese in einer Art Wechselwirkung stehen.
- c) und d) Vor dem Hintergrund von Lehr-/Lerntheorien werden Medien und Unterrichtsmethoden in der Praxis eingesetzt und angewandt. Die daraus gewonnen Erkenntnisse führen wiederum zu einer Überprüfung, Modifizierung und/oder Neuentwicklung von Lehr- und Lerntheorien, die sich ggf. auf die Bildungspolitik auswirken. Der DBR-Ansatz bietet einen Rahmen für die systematische Realisierung des (Weiter)Entwicklungsprozess.

Eine idealtypische Umsetzung des DBR-Prozess unter Einbeziehung der Ebenen wird im Folgenden beschrieben: Der Anlass für den Einsatz von DBR ist meist auf den unteren Ebenen, den Ebenen der Instruktionstechnologien und Unterrichtsmethoden sowie der Lehr-Lernpraxis anzusiedeln. Er ergibt sich aus unterschiedlichsten Herausforderungen und Problemen, mit denen Praktizierende in Lehr-Lernsituationen konfrontiert sind. Die Probleme und Herausforderungen äußern sich meist in einer mangelnden Passung zwischen Zielen der Bildungspolitik und der Erreichbarkeit dieser Ziele. Beispielsweise sind keine adäquaten Instruktionstechnologien vorhanden, um die Inhalte von Curricula effizient zu erreichen. Diese mangelnde Passung kann von Lehrenden und Lernenden als Problem identifiziert werden. Ein Problem in der Lehr-/Lernpraxis gibt schließlich den Anlass für den DBR-Prozess⁴⁹ (Phase 1). Es gilt, eine Lösung für das Problem zu entwickeln (Phase 2) und dieses in mehreren Design-Zyklen bzw. Iterationszyklen (T. Anderson & Shattuck, 2012) zu evaluieren und zu modifizieren (Phase 3). Resultieren sollen daraus sowohl ein theoretischer Output als auch ein praktischer Output (Phase 4). Im Idealfall resultieren aus dem DBR-Prozess innovative Lösungen für ein Problem in der Bildungspraxis, beispiels-

⁴⁹ Der DBR-Prozess wird ausführlich in den Kapiteln 3.1.2 und 8.2.1 erläutert.

weise neu entwickelte Lernmaterialien in Form einer Lernsoftware (Resultat für die Ebene der Instruktionstheorien und Unterrichtsmethoden sowie der Lehr-/Lernpraxis). Auf der Ebene der Lehr- und Lerntheorien werden auf der Grundlage der Ergebnisse des DBR-Prozess neue Theorien generiert oder bestehende modifiziert. Diese tragen ggf. zur Optimierung von Curricula, Zertifikaten, Bildungsstandards und –normen bei.

Somit gilt für DBR einerseits, konkrete Verbesserungen für die Praxis und die Entfaltung innovativer Potenziale im Bildungsalltag zu ermöglichen (praktischer Output). Andererseits sollen der Lösungsprozess und die Lösung hinsichtlich ihres Transferpotenzials überprüft werden, um schließlich einen theoretischen Output in Form von kontextualisierten Theorien zu erzeugen. Edelson (2002) differenziert und kategorisiert die kontextualisierten Theorien in *Domain Theories*, *Design Frameworks* und *Design Methodologies* (Edelson, 2002, S. 113).

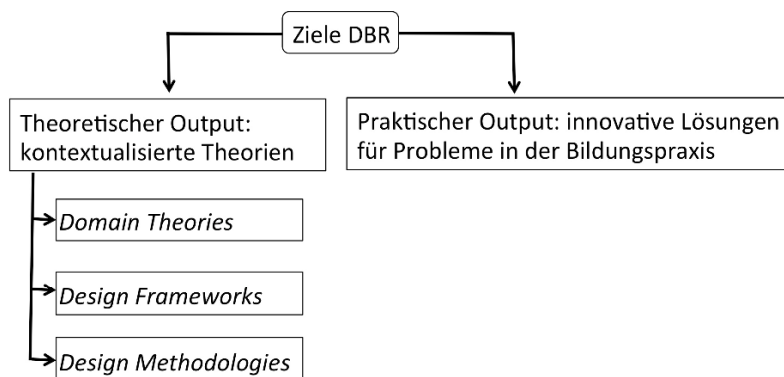


Abbildung 9: Ziele von DBR (in Anlehnung an Edelson, 2002)

Domain Theories (bereichsspezifische Theorien) werden im Kontext von DBR-Prozessen auf der Grundlage von Problemanalysen entwickelt. Prozesse und Ergebnisse, die in Zusammenhang mit bestimmten Interventionen hervorgerufen werden, werden reflektiert und ggf. in eine Theorie transferiert. In ihnen kann thematisiert sein, wie Lernende lernen, wie Lehrende lehren sowie wie Lehr-Lernumgebungen Lehren und Lernen beeinflussen. *Domain Theories* können somit einen natürlichen Mehrwert von Problemanalysen der Lehr-Lernpraxis darstellen (Edelson, 2002, S. 113). Edelson (2002) nennt als Beispiel für eine *Domain Theory* die Theorie über Herausforderungen des Portfolio-Einsatzes in der Schule. In dieser Theorie werden Schwierigkeiten und Probleme berücksichtigt, die Schüler und Schülerinnen haben wenn sie einen Prozess systematisch

beobachten, dokumentieren und reflektieren sollen. Bereichsspezifische Theorien können die theoretische Grundlage sowohl für die Praxis als auch die Forschung bilden. Sie dienen der Optimierung des Lehr-Lernprozesses und beziehen sich auf Lehren, Lernen sowie Lernumgebungen und deren Einfluss auf Lehren und Lernen. Zudem können sie in der Entwicklung von Curricula und Zertifikaten berücksichtigt werden sowie Standards prägen. Edelson erachtet die Entwicklung von bereichsspezifischen Theorien allerdings nicht als Bedingung eines erfolgreichen DBR-Prozesses an. Die Ziele zu formulieren ist zwar Bestandteil des DBR-Prozesses, die Erreichung der Ziele hängt allerdings von der Datengrundlage ab.

Design Frameworks sind generalisierte Design-Lösungen, die sowohl in den Theorien für die Lehr-Lernpraxis als auch im DBR-Prozess selbst – beispielsweise für die Entwicklung von Instruktionstechnologien – berücksichtigt werden können. Im Gegensatz zu den *Domain Theories* sind *Design Frameworks* präskriptiv: sie geben einen Weg bzw. eine Methode für die Bildung und Überprüfung von Hypothesen vor. Hierbei gilt es, kohärente Leitlinien für die Gestaltung von Lernumgebungen (auch „didaktische Szenarien“) zu formulieren.

Design Methodologies sind ebenfalls der Ebene der Lehr-Lerntheorien zuzuordnen und beinhalten die Beschreibung eines verallgemeinerten Ablaufs von Gestaltungsprozessen. Wie auch *Design Frameworks* sind *Design Methodologies* präskriptiv. Sie beschreiben den Lösungsprozess, die Expertise-Anforderungen sowie die Rollen, in denen die Expertise-Anforderungen erfüllt werden. Somit werden in einer *Design Methodology* idealerweise Aufgaben, Ziele, Prozesse und Akteure für jeden Entwicklungs-Schritt bestimmt. Das Ergebnis, das durch den Einsatz einer *Design Methodology* erreicht wird, kann wiederum eine Instruktionstechnologie sein, die in der Praxis eingesetzt wird.

Die Ziele des DBR-Ansatzes fokussieren somit nicht Erklärungen im statistischen Sinn, sondern das Verstehen eines Vorganges (Bruner, 1994; zitiert nach Gravenmeijer & Cobb, 2006, S. 47), welches die empirische Grundlage dafür liefert, wie Interventionen wirken. Zu berücksichtigen ist, dass die Erreichung der Ziele von den Ergebnissen des DBR-Prozesses abhängig ist und die Zielerreichung nicht notwendiger Bestandteil eines erfolgreichen DBR-Prozesses ist. Zusammengefasst sind in einem klassischen DBR-Prozess vier Ebenen involviert. Mit dem DBR-Prozess soll ein praktischer (innovative Lösungen für die Bildungspraxis) und ein theoretischer Output erzeugt werden. Der theoretische Output ist in *Domain Theories*, *Design Frameworks* und *Design Methodologies* differenzierbar. *Domain Theories* fokussieren den Lehr-Lernprozess, *Design Frameworks* die Gestaltung dieser und *Design Methodologies* den Gestaltungsprozess, in dem sowohl Erkenntnisse der *Domain Theories* als auch des *Design*

Frameworks berücksichtigt werden. Der theoretische Output kann auf mehreren Ebenen des DBR verortet sein.

3.1.1 Merkmale von DBR

Auf der Grundlage der des Begriffs-Diskurses und bisheriger DBR zeichnet sich der DBR im Wesentlichen durch sechs Charakteristiken aus:

Charakteristik	Beschreibung
Iterativ ⁵⁰	DBR ist iterativ, indem mehrere Phasen mit Mikro-Design-Zyklen bestehend aus Gestaltung, Durchführung, Analyse und Re-Design in realen Kontexten durchlaufen werden (z.B. Akker u. a., 2006; Bereiter, 2002; Cobb u. a., 2003; Edelson, 2002; Reeves, 2006; Reinmann, 2005).
Interventionsorientiert	Die Forschung besteht aus Interventionen in einem realen Umfeld (z.B. Akker u. a., 2006; Bereiter, 2002; Cobb u. a., 2003; Edelson, 2002; Reeves, 2006; Reinmann, 2005).
Kollaborativ	Wissenschaftler/Wissenschaftlerinnen und Praktiker/Praktikerinnen kooperieren, indem Praktiker/Praktikerinnen den Anlass für DBF identifizieren sowie das Feld zur Verfügung stellen und Forschenden den Lösungsprozess wissenschaftlich begleiten und reflektieren (vgl. z. B. T. Anderson & Shattuck, 2012; Middleton u. a., 2008; Reeves, 2006; Sari & Herrington, 2013; Wang & Hannafin, 2005).
Nutzenorientiert	Da DBR Interventionen zum Ziel hat, durch welche die Wirkungsmechanismen zur Steigerung der Lerneffektivität identifiziert und Innovationen realisiert werden können (vgl. auch Kap. 3.1 zu Ziele des DBR), ist der Mehrwert eines Designs hinsichtlich der Praktikabilität in realen Kontexten zu messen. Der DBR-Prozess orientiert sich an dem Problem, welches es für die Praxis zu lösen gilt, liefert damit aber sowohl für die Praxis als auch für den theoretischen Diskurs einen Mehrwert (vgl. z.B. Akker u. a., 2006; T. Anderson & Shattuck, 2012; Bereiter, 2002; Edelson, 2002; Reeves, 2006; Wang & Hannafin, 2005).
Theoriegeleitet	Der DBR-Prozess wird auf der Grundlage und unter Berücksichtigung aktueller Theorien und Design-Prinzipien durchgeführt (vgl. z. B. Akker u. a., 2006; Cobb u. a., 2003; Edelson, 2002; Reeves, 2006; Sari & Herrington, 2013).

⁵⁰ Für eine genaue Darstellung s. Kap. 3.1.2 zum Ablauf von DBR.

Integrativ	Im DBR-Prozess wird sich verschiedener wissenschaftlicher Methoden bedient, um die Aussagekraft einer Studie zu maximieren. Die Methoden werden kontext- und phasenabhängig ausgewählt und unter Berücksichtigung der wissenschaftlichen Gütekriterien eingesetzt (vgl. z. B. T. Anderson & Shattuck, 2012; Cobb u. a., 2003; Wang & Hannafin, 2005).

Tabelle 7: Charakteristiken von DBR

Wie auch das folgende abschließende Zitat zeigt, zeichnet sich DBR schlussfolgernd weniger durch eine bestimmte Methode aus, sondern durch das Ziel, innovative Lösungen für die Praxis zu liefern, für deren Erreichung sich verschiedener wissenschaftlicher Methoden bedient wird.

„[Design-based research] is not defined by Methodology. [...] What defines design research is its purpose: sustained innovative development“ (Bereiter, 2002, S. 330).

3.1.2 Ablauf von DBR

Auch wenn bisher keine Einheitlichkeit in der Darstellung des Ablaufs von DBR besteht, scheinen sich die Experten und Expertinnen in grundlegenden Aspekten einig zu sein. DBR ist iterativ und prozessorientiert (Akker u. a., 2006; Cobb u. a., 2003; Reeves, 2006; The Design-Based Research Collective, 2003). Ein Black-Box-Modell von Input-Output-Messungen soll vermieden werden; vielmehr liegt der Fokus auf dem Verstehen und der Verbesserung von Interventionen (Akker u. a., 2006). Der DBR-Prozess besteht aus mehreren (in den meisten Beschreibungen vier) Phasen (Cobb u. a., 2003; Middleton u. a., 2008; Reeves, 2006; Reinmann, 2005). Bisher besteht jedoch keine Einheitlichkeit in der Benennung der einzelnen Prozessphasen (vgl. z. B. Edelson, 2002; Middleton u. a., 2008; Reeves, 2006; Reinmann, 2005), auch wenn die Erläuterungen mitunter nur unwesentlich voneinander abweichen. In den meisten Darstellungen besteht eine der (vier) Phasen wiederum aus Mikro-Design-Zyklen, die mehrfach wiederholt werden können (Gravenmeijer & Cobb, 2006; Reeves, 2006; Reinmann, 2005). Reeves erläutert den Ablauf von DBR am Beispiel einer Technologieentwicklung (2006, S. 96)⁵¹:

⁵¹ Da die Beschreibung der vier Phasen in der Literatur nur unwesentlich voneinander abweichen, wird auf eine differenzierte Darstellung unterschiedlicher Autoren und Autorinnen verzichtet. Zur Veranschaulichung des DBR-Ansatzes wird hier das mehrfach zitierte Modell von Reeves herangezogen, da Reeves den Ablauf anhand einer Technologieentwicklung verdeutlicht und der Untersuchungsgegenstand dieser Arbeit ebenfalls eine technische Entwicklung ist.

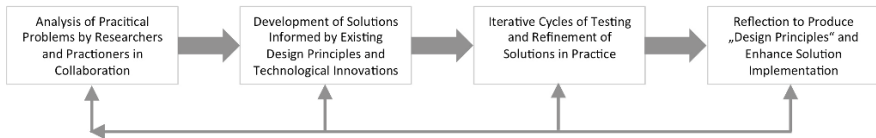


Abbildung 10: DBR-Prozess nach Reeves (2006, 96)

Die erste Phase besteht aus der Problemanalyse, auf die in der zweiten Phase die Entwicklung von Lösungen unter Berücksichtigung von Design-Prinzipien und technischen Entwicklungen folgt. In der dritten Phase werden Zyklen von Testung und Modifikation durchlaufen, um schließlich Design-Prinzipien und Lösungsumsetzungen zu entwickeln.

Eine ausführliche Erläuterung der vier Phasen wird vom Autor nicht vorgenommen. Middleton u. a. (2008) greifen das Modell von Reeves auf und konkretisieren die einzelnen Phasen. Neben Middleton u. a. weisen weitere Prozessdarstellungen starke Parallelen zu dem Modell von Reeves auf (vgl. Edelson, 2002; Gravenmeijer & Cobb, 2006). Zur Erläuterung der Phasen werden die Ergänzungen weiterer Autoren und Autorinnen herangezogen. Die Bezeichnungen der einzelnen Phasen werden im Folgenden in deutscher Sprache verkürzt betitelt.

Phase 1: Problemanalyse (*Analysis of Practical Problems by Researchers and Practitioners in Collaboration*): Der Anlass für eine Untersuchung wird zumeist aus der Praxis geliefert. In der ersten Phase wird ein Problem bzw. ein Forschungsdesiderat erkannt, welches im Idealfall zusammen mit Wissenschaftlern und Wissenschaftlerinnen konkretisiert wird (Reeves, 2006, S. 97). Es wird ein Anliegen formuliert, dessen Eigenschaft und Hintergrund exploriert werden soll. Auf der Grundlage eines systematischen Literatur-Reviews und der Berücksichtigung aktueller Forschungsergebnisse wird die Erforschung des Anliegens legitimiert und deren Relevanz dargestellt. Dies beinhaltet sowohl Formulierung der praktischen als auch theoretische Relevanz der Untersuchung (Jan Herrington, McKenney, Reeves & Oliver, 2007, S. 4094).

Phase 2: Konzeption (*Development of Solutions Informed by Existing Design Principles and Technological Innovations*): Sobald eine klare Forschungsfrage formuliert wurde, wird ein theoretisches Konzept⁵² entwickelt. Das Konzept bedient mutmaßlich das in der Praxis identifizierte Problem bzw. die Forschungsfrage. Um die Komplexität eines realen Settings zu durchdringen und „evaluierbar“ zu machen, werden auf Basis der aktuellen Theorien sowie des darauf aufbauenden Konzeptes einerseits Kriterien entwickelt, welche für die

⁵² Das theoretische Konzept kann wiederum die Grundlage für das Ziel, ein *Design Framework* zu entwickeln (vgl. Kap. 3.1) sein.

Evaluation heranzuziehen sind. Andererseits werden wissenschaftliche Methoden identifiziert, mit Hilfe derer das Konzept evaluiert werden soll. Zudem wird ein Testdesign konzipiert, welches die Überprüfung des Konzeptes zum Gegenstand hat. Das Konzept kann anhand der Ergebnisse auf Veränderungen bzw. Weiterentwicklungen hin überprüft werden. Es muss also „testbar“ gegenüber vorheriger, das Artefakt bestimmenden Merkmalen, sein. Im Idealfall geschieht dies durch Kontrollgruppen mit einer randomisierten Zuordnung (Middleton u. a., 2008, S. 28).

Phase 3: Konzeptevaluation (*Iterative Cycles of Testing and Refinement of Solutions in Practice*): Das Konzept wird schließlich in mehreren Zyklen – auch Design-Zyklen genannt (Reinmann, 2005) – überprüft.

„Design research involves the continuous iterative cycle of design, enactment, implementation, analysis, and redesign“ (The Design-Based Research Collective, 2003, S. 5).

Die Bezeichnungen und einzelnen Schritte der Zyklen sind nicht einheitlich, weichen aber inhaltlich nur wenig voneinander ab. Da die Mikro-Zyklen in den meisten Darstellungen aus vier Schritten bestehen, wird auch im weiteren Verlauf auf diese Darstellung zurückgegriffen. Die Zyklen beinhalten a. Gestaltung, b. Durchführung, c. Analyse und d. Re-Design (Gravenmeijer & Cobb, 2006, S. vgl. auch; Middleton u. a., 2008; Reinmann, 2005).

a. Gestaltung

Die Gestaltungsphase läuft in drei Teilschritten ab und wird sowohl auf der praktischen als auch auf der methodischen Ebene umgesetzt⁵³.

a1: Entwicklung und Definition des Evaluationsgegenstands

Auf der praktischen Ebene wird der Evaluationsgegenstand unter Berücksichtigung aktueller theoretischer Erkenntnisse entwickelt und definiert. Ggf. wird eine Literaturrecherche durchgeführt, um den aktuellen Stand der Forschung in die Entwicklung einbeziehen zu können. Sind keine passenden Theorien für den eigenen Ansatz antizipierbar, werden vorwissenschaftliche Erfahrungen eingebracht. Bei dem Evaluationsgegenstand handelt es sich beispielsweise um ein Lernszenario oder eine technische Entwicklung.

a2: Formulierung der Forschungsfragen

Vor dem Einsatz und der Testung werden/wird die Forschungsfrage(n) formuliert und operationalisiert. Die Forschungsfragen können explorativen oder aber auch hypothesengeleiteten Charakter haben.

⁵³ Einige Schritte gleichen denen der ersten Phasen des gesamten Design-Prozesses, sie werden hier allerdings auf der Mikroebene umgesetzt.

a3: Methodenwahl und Bestimmung des Untersuchungskontextes

Auf der methodischen Ebene wird das Forschungsdesign entwickelt. Es werden Evaluationskriterien, Untersuchungskontext, Erhebungsmethoden sowie Probanden und Probandinnen bestimmt.

b. Durchführung

In der Durchführungsphase wird das Design in der Praxis erprobt und es wird die Interaktion zwischen Medien, Lehrenden und Lernenden systematisch erfasst (Reinmann, 2005).

c. Analyse⁵⁴

Die Analyse ist in zwei Schritte zu unterteilen.

c1: Datenauswertung

Im Rahmen des ersten Analyseschritts werden die Ergebnisse ausgewertet. Für die Auswertung werden der Erhebungsmethode entsprechend adäquate Auswertungsmethoden und -instrumente herangezogen.

c2: Ergebnisdarstellung und -interpretation

Die in der Durchführungsphase gewonnenen Daten werden dargestellt und unter Berücksichtigung der wissenschaftlichen Gütekriterien interpretiert. Daraufhin werden Modifizierungsmöglichkeiten formuliert.

d. Re-Design

Die Modifizierungsideen werden in der Phase des Re-Designs umgesetzt.

Die Zyklen a. und d. können sich mitunter überschneiden: Die Modifizierungen auf der Grundlage der vorherigen Phasen der Design-Zyklen können gleichzeitig auch als Konzeptentwurf dienen, den es dann in der zweiten Phase zu überprüfen gilt.

Die Design-Zyklen können beliebig oft durchlaufen werden.

Phase 4: Reflexion (*Reflection to Produce „Design Principles“ and Enhance Solutions Implementation*): Werden keine weiteren Zyklen durchlaufen, erfolgt nach der Durchführung der letzten Phase im letzten Zyklus eine Interpretationsphase. Deren Ziel ist die Entwicklung kontextualisierter Theorien, abgeleitet aus den empirischen Ergebnissen (Edelson, 2002 und vgl. Kap. 3.1). Anhand der Kriterien des zu Beginn formulierten Konzeptes kann es hinsichtlich seiner Praxistauglichkeit auf der Grundlage der Ergebnisse reflektiert werden. Die folgende Abbildung stellt den DBR-Prozess mit drei Design-Zyklen schematisch dar:

⁵⁴ Die Bezeichnung für die vierte Phase nach Reinmann (2005) „Überprüfung“ wird im Folgenden mit dem Begriff „Analyse“ betitelt. Dies soll stärker betonen, dass für die Überprüfung eines Designs die Analyse die ausschlaggebenden Hinweise für den nachfolgenden Schritt, das Re-Design, liefern.

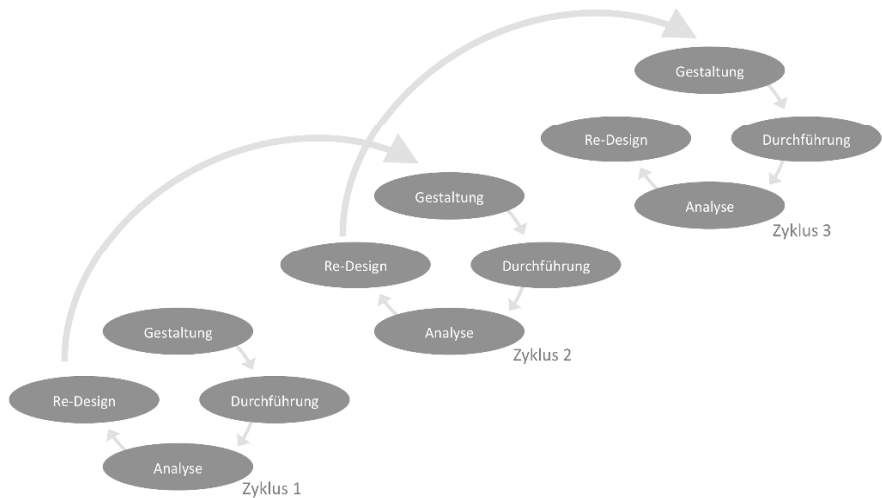


Abbildung 11: Schematischer Ablauf DBR-Zyklen

Zusammenfassend sind die Mikro-Zyklen von Gestaltung, Durchführung, Analyse und Re-Design ein Bestandteil des iterativen Prozesses, der letztlich einen formativen, prozess- und zielorientierten Charakter aufweist und für dessen Zielerreichung sich wissenschaftlicher Methoden bedient wird. Resultat sind ein praktischer (innovative Lösungen für die Bildungspraxis) sowie theoretischer (kontextualisierte Theorien) Output.

3.1.3 Forschungsstrategisches und -methodisches Vorgehen

Die Charakteristiken und der Ablauf von DBR geben zwar Aufschluss darüber *was* zu machen ist, selten jedoch wird das *Wie* erläutert. Um zu aussagekräftigen Forschungsergebnissen zu gelangen und über den einfachen Design-Prozess hinausgehen zu können, sind im DBR-Prozess folgende forschungsstrategische und -methodische Anforderungen zu berücksichtigen (angelehnt an Edelson, 2002):

- a) **Forschungsbasierte und -orientierte Vorgehensweise:** Um brauchbare Forschungsergebnisse zu erzielen, muss das Forschungsprogramm auf vorheriger Forschung aufbauen und von Forschungszielen begleitet werden.
- b) **Systematische Dokumentation:** Im Vergleich zu beispielsweise den Ingenieurwissenschaften (vgl. Kap. 3.1.4) ist in erziehungswissenschaftlichen Designprozessen die systematische Dokumentation nicht üblich. Allerdings ist gerade die systematische Dokumentation der einzige Zu-

gang, um aus einem Designprozess wissenschaftliche Ergebnisse zu erzielen. Hierzu zählen die Dokumentation der theoretischen Standpunkte der Entwickler und Entwicklerinnen, der Problemanalyse, des Lösungsprozesses und des Designprozesses.

- c) **Formativer Charakter:** Die Evaluation des Designprozesses dient der Identifizierung von Unzulänglichkeiten, die in der weiteren Entwicklung berücksichtigt werden können bzw. müssen.
- d) **Generalisierung:** Das finale Element des Design-Based Research ist der Generalisierungsprozess. In diesem versuchen die Wissenschaftler und Wissenschaftlerinnen die Ergebnisse der Problemanalyse, des Lösungsprozesses sowie des Designprozesses auf andere Kontexte zu übertragen. Im Rahmen der Generalisierungen wird somit das Ziel verfolgt, auf einer Meta-Ebene kontextualisierte Theorien zu entwickeln, die in die *Domain Theories*, das *Design Frameworks* und die *Design Methodologies* einfließen können.

In dieser Vorgehensweise spiegeln sich die Charakteristiken von DBR wieder. Beispielsweise impliziert er Anspruch, DBR müsse auf vorheriger Forschung aufbauen (a) die Berücksichtigung aktueller Theorien und somit die Theoriegeleitetheit. Durch die systematische Dokumentation (b) wird die intersubjektive Nachvollziehbarkeit realisiert, die wiederum ein Gütekriterium der qualitativen Forschung darstellt (vgl. Kap. 8.2.5). Damit kommt der integrative Charakter zum Ausdruck. Die Attribute formativ und iterativ (c) stellen sowohl eine Charakteristik als auch das Vorgehen dar. Die Generalisierung (d) kommt in der Charakteristik der Nutzenorientierung zum Ausdruck.

Deutlich wird dadurch, dass sich DBR an klassischen Forschungstraditionen orientiert. Der Frage, wie DBR in den klassischen Forschungstraditionen zu verorten ist, wird im folgenden Kapitel nachgegangen.

3.1.4 Der DBR-Ansatz im wissenschaftlichen Kontext

Wie bereits erwähnt, zeichnet den Design-Based Research-Ansatz keine bestimmte Methode aus. Daher lassen sich ebenso wenig methodologische Vorgehensweisen definieren, die in der Wissenschaftlichen Praxis allgemein anerkannt sind und das Vorgehen bestimmen. Wie in den folgenden Kapiteln erläutert wird, erfüllt der Ansatz dennoch die Ansprüche an ein wissenschaftliches Vorgehen (vgl. z. B. Jan Herrington u. a., 2007; Wang & Hannafin, 2005), indem diverse Ansätze aus den Forschungstraditionen integriert werden.

3.1.4.1 Verortung des DBR im Kontext der Wissenschaftstraditionen

Grob lassen sich die unterschiedlichen Forschungstraditionen in geisteswissenschaftliche, naturwissenschaftliche und ingenieurwissenschaftliche Forschung

differenzieren. Primär lehnt sich DBR an den Ingenieurwissenschaften an. Wie aber durch die folgende Kurzbeschreibung der Forschungstraditionen deutlich wird, weist DBR Schnittstellen zu allen drei Bereichen auf.

Der geisteswissenschaftliche Ansatz ist der älteste im Kontext der Erziehungswissenschaften. Beschrieben wird er als „original investigation undertaken in order to gain knowledge and understanding; scholarship; the intervention and generation of ideas [...] where these lead to new or substantially improved insights“ (Higher Education Funding Council, 1999, S. 4) (zitiert nach Burkhardt & Schoenfeld, 2003, S. 5). Laut Burkhardt und Schoenfeld (2003) sind Indikatoren für die Qualität geisteswissenschaftlicher Forschung die Plausibilität, die interne Konsistenz und die Passung in bisherige Kenntnisse und Theorien. Allerdings wurde deutlich, dass einige der entwickelten geisteswissenschaftlichen Theorien und Ideen in der Praxis nicht wirksam waren und es an einer empirischen Basis mangelte – einerseits an empirischen Erkenntnissen und andererseits an empirischen Methoden. Dies führte zu einer Suche nach einer – und letztlich stärkeren – Integration evidenzbasierter Erziehungswissenschaft. So wurden zunehmend Methoden aus den Naturwissenschaften adaptiert (Burkhardt & Schoenfeld, 2003, S. 5).

Der naturwissenschaftliche Ansatz ist vom Erkenntnisinteresse getrieben – mit Hilfe von Erklärung und Analysen von Phänomenen und von Entwicklungen von Modellen, welche die Phänomene erklären. Ziel ist, beobachtete Phänomene mit empirischen Methoden und Untersuchungen zu erklären. Die naturwissenschaftliche Forschung selbst fokussiert somit nicht primär die Generierung praktischer Lösungen, was hingegen in den Ingenieurwissenschaften im Vordergrund steht.

Die Ingenieurwissenschaften stehen in direkter Beziehung zur Praxis und verfolgen neben dem Anspruch, die Welt zu erklären vorwiegend das Ziel, Arbeitsprozesse zu unterstützen, indem systematisch qualitativ hochwertige auf der aktuellen Theorie basierende Lösungen für praktische Probleme entwickelt werden (Burkhardt & Schoenfeld, 2003, S. 5). In den Ingenieurwissenschaften wird auf den bisherigen Erkenntnissen und Erfahrungen (auch aus anderen Disziplinen) aufgebaut. Die Ingenieurwissenschaften werden beschrieben als „the use of existing knowledge in experimental development to produce new or substantially improved materials, devices, products and processes, including design and construction“ (Higher Education Funding Council, 1999, S. 4) (zitiert nach Burkhardt & Schoenfeld, 2003, S. 5). Im Entwicklungs- und Evaluationsprozess werden Designentwürfe mit empirischer Testung kombiniert, um Werkzeuge und Prozessabläufe zu entwickeln bzw. zu optimieren. Ingenieure und Ingenieurinnen bearbeiten verschiedene Problemstellungen, die auf der Basis aktueller Theorien und Forschungsergebnisse gelöst werden sollen. Neben einem praxis-

tauglichen Output werden neue Design- und Testmethoden entwickelt. Hier gehen die Ansätze der Ingenieurwissenschaften einen Schritt weiter als es im DBR-Ansatz vorgesehen ist: Ingenieurwissenschaftler und -wissenschaftlerinnen erforschen das entwickelte Produkt hinsichtlich seiner Disseminations- und Transfermöglichkeiten (Wilhelm, Tobias, Waltner, Hopf & Wiesner, 2012, S. 31), was im DBR-Prozess zwar in der vierten Phase ebenfalls realisiert werden kann, aber keine Bedingung ist.

Als Vorbild für DBR-Ansätze fungiert in erster Linie die Ingenieurwissenschaft. In der erziehungswissenschaftlichen Forschung wird der ingenieurwissenschaftliche Ansatz allerdings wenig berücksichtigt und teilweise unterbewertet. Als gängig und „veröffentlichungswürdig“ gelten vornehmlich erklärende Ansätze („*insight research*“), wie sie die Geistes- und Naturwissenschaften postulieren. In der Erziehungswissenschaft sind Methoden der Ingenieurwissenschaften wenig vertreten. Dabei können gerade Untersuchungen, in denen grundlagen- und anwendungsorientierte Forschung verknüpft werden, Innovationen hervorbringen. Bei dem Problemlöseprozess entstehen ggf. neue Fragen, die beantwortet werden müssen bzw. können, was zu neuen Erkenntnissen und einem Wissenszuwachs führen kann. Auch wenn erklärungsorientierte Forschung eine wichtige Grundlage für die Weiterentwicklung der Praxis sein kann, liefert sie keine Lösungen, die direkt in die Praxis implementiert werden können. Um das zu erreichen, sind sowohl forschungsbasierte Entwicklungen als auch deren Evaluation in der Praxis notwendig. Bei DBR handelt es sich um einen Ansatz, in dem beides verknüpft wird.

Es wird kontrovers diskutiert, inwieweit der DBR-Ansatz den Ansprüchen des wissenschaftlichen Arbeitens sowie den Gütekriterien der Objektivität, Reliabilität und Validität genügt bzw. überhaupt genügen muss. Fischer u. a. postulieren, dass Design-Based Research keines der vom National Council veröffentlichten Prinzipien zur Wissenschaftlichkeit verletzt (Fischer, Bouillion & Mandl, 2003, S. 156 f). Des Weiteren ist beispielsweise Edelson (2002) der Meinung, dass Design-Based Research keine statistisch abgesicherten Daten im Sinne empirischer Forschung liefert bzw. liefern muss. An dieser Stelle werden zwei Unterschiede zwischen Design-Based Research und empirischer Forschung deutlich:

Einerseits stehen die Einhaltung der Gütekriterien und die Hypothesenüberprüfung nicht im Vordergrund. Vielmehr sind die Messkriterien für Design-Based Research *Neuheit, Nützlichkeit und Innovation* (Reinmann, 2005). Zum anderen besteht die Stärke des Ansatzes nicht in der statistischen Validierung, sondern im explorativen Charakter und dem Fundament aus kontextuellen Erfahrungen (Edelson, 2002, S. 117 f). DBR ist daher nicht inkompatibel mit traditioneller Outcome-orientierter empirischer Forschung. Der Ansatz liefert Theorien,

die ihrerseits wiederum statistisch überprüft werden können. Zudem begegnet sie den Gefahren einer „Überfokussierung auf empirische Forschung“ (Edelson, 2002, S. 118), die einerseits dazu führen kann, dass Theorien empirisch überprüft und aufgrund der Ergebnisse als nicht tauglich eingestuft werden, obwohl sie zum Zeitpunkt der Evaluation noch nicht voll entwickelt waren. Andererseits kann eine Fokussierung dazu führen, dass zu sehr die Evaluation und weniger die Entwicklung und Innovation im Vordergrund steht (Edelson, 2002, S. 118). Somit ist der DBR-Ansatz eine hilfreiche Ergänzung, um die Lücke mangelnder Innovationskraft in den Bildungswissenschaften (vgl. Reinmann, 2005) zu schließen.

3.1.4.2 DBR in Abgrenzung zu weiteren Ansätzen in der Lehr-Lernforschung

Wie bereits in vorangegangenen Kapiteln beschrieben, kann sich im Kontext von DBR unterschiedlicher wissenschaftlicher Methoden bedient werden. Naheliegender ist demnach, dass DBR Parallelen zu weiteren wissenschaftlichen Disziplinen und Methoden aufweist, sich aber auch in bestimmten Aspekten eindeutig abgrenzen lässt. Um die Besonderheiten und Gemeinsamkeiten herauszuarbeiten, wird DBR im Folgenden mit (teil)verwandten Methoden verglichen – allerdings ohne den Anspruch eines flächendeckenden Vergleichs, in dem jegliche Methoden, die Parallelen zum DBR aufweisen, angeführt werden.

Im Vergleich zur experimentellen Forschung (wie z. B. Aptitude-Treatment-Interaction) haben die Interventionen im DBR-Ansatz eher einen holistischen Charakter. Die Interventionen bestehen aus der Interaktion zwischen Methoden, Medien, Materialien, Lehrenden und Lernenden. Eine Intervention ist demnach das Produkt des Kontextes, in dem sie durchgeführt wird (Reinmann, 2005, S. 63). Für ein experimentelles Design ist dieses Vorgehen aufgrund der meist vielzähligen Interaktionsmöglichkeiten nur bedingt geeignet. Denn die experimentelle Forschung variiert einzelne Bedingungen in einem kontrollierbaren Setting und beobachtet deren Auswirkungen. Beispielsweise wird in der Aptitude-Treatment-Interaktion der Frage nachgegangen, welche individuell gegebenen Lernvoraussetzungen bei Anwendung unterschiedlicher Lernmethoden zu bestimmten Lernzielen führen (Hasebrook, 2006, S. 20). Wie auch der Begriff aussagt handelt es sich um die Untersuchung von Wechselwirkung anhand der Variation von Variablen. Während sich die experimentelle Forschung über die klassischen Gütekriterien legitimiert, wird die Aussagekraft der DBR-Forschung auf Grundlage der Kriterien Neuheit, Nützlichkeit und nachhaltige Innovation bewertet.

Wie auch in der Evaluationsforschung werden im DBR multiple Methoden eingesetzt, eine iterative Vorgehensweise wie beim DBR ist dabei nicht vorgesehen. Die Evaluation hat zum Ziel, Konzepte, Designs, die Umsetzung und den

Nutzen sozialer Interventionsprogramme unter systematischer Anwendung sozialwissenschaftlicher Forschungsmethoden zu beurteilen (Wottawa, 2006, S. 650). Der Unterschied zur Evaluationsforschung besteht allerdings darin, dass es bei DBR nicht nur um die Überprüfung und Perfektionierung eines Produktes, sondern auch um ein theoretisches Verständnis von Lehren und Lernen sowie um die Entwicklung von Theorien, die nicht nur auf den Untersuchungskontext zutreffen und auf weitere Kontexte zu übertragen sind, geht. Auch die in der Evaluationsforschung häufig eingesetzte Methode des formative Assessments (vgl. auch Kap. 2.1) unterscheidet sich in ähnlicher Hinsicht, denn formative Assessment hat zum Ziel, Theorien zu testen und nicht, neue Theorien zu entwickeln, wie es im DBR vorgesehen ist (Barab & Squire, 2004). In der Theorie des Instruktionsdesign geht es zwar auch um die Planung, Entwicklung und Evaluation von Lernumgebungen und -materialien doch grenzt diese sich durch seine stark lern- und kognitionspsychologische Begründung ab. Zudem sind im Rahmen des Instruktionsdesigns experimentelle Studien mit dem Fokus auf das lernende Individuum (Briggs & Gagné, 1974) vorherrschend.

Zu der insbesondere in den 1970er Jahren populär gewordenen Aktionsforschung (Action Research) unterscheidet sich DBR durch die Berücksichtigung theoretischer Erkenntnisse und der Einbeziehung von drei Personenkreisen. In der Aktionsforschung nimmt eine Person sowohl die Rolle der/des Lehrenden als auch der Forscherin/des Forschers ein. In DBR sind Forschende, Lehrende sowie Lernende involviert. Des Weiteren grenzt sich die Aktionsforschung dadurch ab, dass sie nicht die Generalisierung der Ergebnisse bzw. den Transfer auf andere Kontexte verfolgt. Vielmehr hat die Aktionsforschung die Generierung von Erkenntnissen und Entwicklungen für die Praxis zum Ziel, die mit Hilfe mehrerer Forschungs- und Entwicklungszyklen gewonnen werden. Die Aktionsforschung fordert damit eine experimentelle Haltung der eigenen Praxis gegenüber (Posch & Zehetmeier, 2010, S. 4). Dazu gehört auch der flexible Einsatz von Methoden, die ggf. im Laufe des Forschungsprozesses modifiziert werden sollen. Durchgeführt wird der Forschungsprozess von Forschenden und Personen, die direkt betroffen sind (z. B. Lehrende im Unterricht). Die Untersuchungspersonen werden über die Forschungsziele informiert und sollen über die Ziele und Auswertungen mitbestimmen können. Ein Leitsatz der Aktionsforschung ist die Gleichberechtigung aller Beteiligten. Dementsprechend stößt die Aktionsforschung an ihre Grenzen, wo kein ausreichendes Problembewusstsein seitens der Untersuchungsteilnehmenden vorhanden ist. Aktionsforschung sei eher für gut gebildete Teilnehmende geeignet (Bortz & Döring, 2006, S. 345). Kritik wird an ihr im Wesentlichen in ihrem in der Praxis schwer aufrecht zu erhaltenden Anspruch der Gleichberechtigung und der möglichen Selbsttäuschung der Wissenschaftler und Wissenschaftlerinnen geübt (Spieß, 1994, S. 4). Der Aktionsforschung wird

vorgeworfen, die Ergebnisse auf gegenwärtige Praxiskontexte zu übertragen, ohne jedoch weitere Vorbedingungen (wie beispielsweise das Vorwissen) und bisherige Kenntnisse zu berücksichtigen. So muss sie sich dem Vorwurf aussetzen, nicht ausreichend wissenschaftlich zu sein. Das Ausblenden bisheriger Erkenntnisse und Theorien und mangelnde Transferbemühungen sind letztlich auch die wesentlichen Aspekte, durch die sich Aktionsforschung und DBR unterscheiden.

Schließlich zeichnet sich DBR durch den hohen Stellenwert des Designs aus und grenzt sich eben auch dadurch von den eben genannten Ansätzen ab. Im DBR wird unmittelbar ein praxistaugliches Design fokussiert. In der klassischen empirischen pädagogischer Forschung geht es vornehmlich um die Generierung neuer Erkenntnisse, aus denen ggf. erst in einem nachfolgenden Schritt praxistaugliche Ideen abgeleitet werden. Um die Bandbreite des Anwendungsfelds von DBR aufzuzeigen, werden im folgenden Kapitel Beispiele aus der Praxis vorgestellt.

3.1.5 DBR in der Anwendung

Anderson und Shattuck (2012) analysierten 47 fachwissenschaftliche Artikel aus den Jahren 2002 bis 2011, in denen Anwendungsbeispiele von DBR im erziehungswissenschaftlichen Kontext erläutert und reflektiert werden. 34% der Artikel behandeln das Thema aus einer philosophischen bzw. erklärenden Perspektive. Die restlichen Artikel (vornehmlich aus den Jahren 2007 bis 2010) legen empirische Ergebnisse, gewonnen aus DBR-Anwendungen, dar (T. Anderson & Shattuck, 2012, S. 16). 73% der Artikel wurden von Autoren und Autorinnen in den USA verfasst, darauf folgen mit nur 5% die Niederlande; von deutschen Autoren und Autorinnen stammen lediglich 2%. In über der Hälfte der Studien, die einen Anwendungsbezug aufzeigen, werden drei oder mehr Iterationszyklen angewendet (T. Anderson & Shattuck, 2012, S. 20). Die meisten Ergebnisse liefern Hinweise auf Verbesserungen des Lernverhaltens, der Lernmotivation bzw. des Lernens selbst. Ungefähr ein Drittel trug zum Verständnis von erziehungswissenschaftlichen Phänomenen bei (T. Anderson & Shattuck, 2012, S. 21). Zusammenfassend postulieren Anderson und Shattuck, dass DBR insbesondere für den Einsatz in oberen Schulstufen mit Technologiebezug attraktiv erscheint (T. Anderson & Shattuck, 2012, S. 24).

Ein Beispiel für die Anwendung von DBR aus dem US-amerikanischen Raum ist die Studie von Squire (2005), in der untersucht wurde, wie Lern-Videospiele konzeptionell und visuell zu gestalten sind, um zur Motivationssteigerung und zu einem Lerneffekt beizutragen. Ausgangslage bzw. das identifizierte Problem war, dass Videospiele als unterhaltsam und herausfordernd, E-Learning-Systeme aber nach wie vor als langweilig bewertet wurden. Videospie-

le, die zum Lernen entwickelt wurden, besaßen nicht die gleiche Attraktivität wie Videospiele, die der Unterhaltung dienen. Um Lernspiele für Lernende attraktiver zu gestalten, ging es nicht nur um das pädagogische Potenzial des Mediums selbst, sondern es galt insbesondere zu untersuchen, welche Faktoren bei Videospielen motivationsfördernd sind. Hierfür sollten mit dem DBR-Ansatz Einblicke in die Motivation, die Adaptionmöglichkeiten als auch die Schule als soziale Organisation bezüglich der Nutzung von Videospielen gewonnen werden. Die Iterationen beinhalten Gedankenexperimente (erste Iteration) sowie Entwicklung, Einsatz und Evaluation mehrerer Prototypen. In der zweiten Iteration wurden die Prototypen mit Studierenden, Lehrenden und Videospielentwickler/-entwicklerinnen getestet, um gruppenspezifische Effekte zu untersuchen. Mit ca. 2000 Versuchsteilnehmenden wurden sowohl ein praktischer (Entwicklung einer möglichst einfachen Steuerung) als auch ein theoretischer (Verständlichkeit verschiedener Spielgenres) Output verfolgt. Die Beobachtungen über die Wirkung und Verständlichkeit unterschiedlicher Genres führte schließlich zu der Erkenntnis, dass einige bestimmte Genres eine größere Nutzer-/Nutzerinnenschaft ansprechen, was es schließlich auch in der Entwicklung von Lernspielen zu berücksichtigen galt. Nachdem ein gebrauchstaugliches Interface entwickelt war, wurde das Spiel in einer dritten Iteration mit Studierenden sowie Professoren und Professorinnen getestet. Eingesetzt wurden dabei typische Usability-Evaluationsmethoden, z. B. Thinking-Aloud-Experimente. Die Ergebnisse zeigen, dass die inhaltliche Ausrichtung hinsichtlich der Aufgabenschwierigkeit noch nicht dem Vorwissen der Zielgruppe entsprach, woraufhin weitere Spiel-Level entwickelt wurden, in denen dieses Vorwissen vermittelt werden sollte. In einer weiteren Iteration wurde das Produkt mit Schülern und Schülerinnen höherer Stufen (z. B. High School) getestet, wobei das „Curriculum“ des Spiels in Zusammenarbeit mit den Lehrenden entwickelt wurde. Für die Evaluation wurden diverse wissenschaftliche Methoden eingesetzt: Qualitative (Feldbeobachtungen, Videoaufzeichnungen, Interviews) und quantitative (Pre-Post-Tests) Methoden. Aus den qualitativen Daten wurde abgeleitet, dass Lernende wesentlich mehr Wert auf konzeptionelle Aspekte (z. B. Steuerung und Spielmechanismen) legen als auf die Qualität von Grafiken. Ein weiterer Befund war, dass Mädchen/Frauen vorerst das Spielen verweigerten aber dann durchschnittlich länger spielten als Jungs/Männer. Abgeleitet wurde daraus, dass sich zwar weibliche Personen im Laufe des Spielens zunehmend auf das Spiel einlassen aber anfangs Hürden bestehen können. Zudem wurde in der Studie der Lerneffekt von Schülern und Schülerinnen, die mit Videospielen lernen sowie Schülern und Schülerinnen aus Klassen, in denen keine Videospiele eingesetzt wurden, verglichen. Der Studie zufolge erzielten die Personen aus der Kontrollgruppe durchschnittlich 20% schlechtere Ergebnisse. Somit scheinen die Autoren und

Autorinnen das Ziel ihrer Studie, motivations- und lernfördernde Aspekte in Lernspiele zu implementieren, erreicht zu haben.

Deutlich wird bei dieser Studie, dass in einzelnen Iterationen je nach Entwicklungsstand unterschiedliche Erhebungsmethoden eingesetzt wurden und die Ergebnisse vorheriger Iterationen in die Weiterentwicklung des Produkts flossen. Der Einsatz wissenschaftlicher empirischer Methoden (Usability-Evaluationsmethoden, Interviews, Befragungen etc.) führt zu validen Ergebnissen und erlaubt die Ableitung von Schlussfolgerungen. An dieser Studie wird die Stärke von DBR deutlich: DBR bietet einen Forschungsrahmen, in dem verschiedene Methoden in unterschiedlichen Phasen kombiniert und integriert werden können um für die Verbesserung der Bildungspraxis aussagekräftige Ergebnisse zu erzielen.

Ein weiteres Beispiel von DBR ist die Entwicklung und Evaluation von der Multi-User-Virtual-Environment (MUVE) „River City“⁵⁵, welche u. a. an der Universität Harvard durchgeführt wurden. Konkret wurde in dem Projekt untersucht, ob die virtuellen Umgebungen reale Experimente simulieren können und einen positiven Effekt auf das Lernen, das Engagement im Lernprozess und das wissenschaftliche Verständnis haben (Nelson, Ketelhut, Clarke, Bowman & Dede, 2005). Für die Untersuchung wurden quantitative Multi-Level-Analysen durchgeführt und Kommentare von Schülern und Schülerinnen sowie Lehrenden ausgewertet (Ketelhut & Nelson, 2010). dafür wurde in einem ersten Zyklus die MUVE mit einem papierbasierten Äquivalent verglichen, indem quantitative Pre- und Post-Befragungen, die Erwartungen von Lehrpersonen sowie der Erfolg von Schülern und Schülerinnen qualitativ erfragt wurden (Nelson u. a., 2005). U. a. zeigen die Erhebungen, dass River City für alle Schüler und Schülerinnen – unabhängig ihrer Fähigkeiten in dem entsprechenden Bereichs und des Geschlechts – einen positiven Einfluss auf das Engagement im Lernprozess hat. Modifizierungsbedarfe wurden hinsichtlich der graphischen/curricularen Aspekte, hinsichtlich der Auswahlmöglichkeiten bzw. Selbstgestaltungsmöglichkeiten und der Interaktionsmöglichkeiten der Avatare festgestellt. Diese wurden in der anschließenden Re-Design-Phase berücksichtigt. Im zweiten Zyklus wurden Fokusgruppen aus Lernenden eingesetzt. Die Ergebnisse des zweiten Zyklus zeigen, dass viel Zeit für die Exploration von River City benötigt wird, bevor mit der Bearbeitung begonnen werden kann. Das Risiko, in der MUVE „verloren zu gehen“, war hoch. Die Lernenden suchten den Zugang zu digitalen Handbüchern der MUVE. Überarbeitet wurden daraufhin die Unterstützungsmöglichkeit zur Orientierung in der MUVE und der Zugang zur – in der virtuellen Umgebung integrierten – Bibliothek. Der dritte Zyklus bestand aus zwei Pilot-Studien, die

⁵⁵ www.riversity.activeworlds.com (zuletzt geprüft am 03.11. 2014)

dem Erhebungsdesign des ersten Zyklus glichen. Dabei wurde lediglich weiterer Modifizierungsbedarf hinsichtlich der Orientierungsmöglichkeiten festgestellt. Auf theoretischer Ebene wurde im vierten Zyklus das Konzept des situierten Lernens und das Konzept des Lernens durch Handeln verglichen. Der Vergleich sollte das Verständnis über die Anwendung der entsprechenden Lerntheorien steigern und somit praktisch die Fähigkeit des Wissenstransfers vom Bildungs- in den Praxiskontext fördern. Auf der praktischen Ebene wurden im vierten Zyklus drei Varianten getestet, in denen unterschiedliche lerntheoretische und -didaktische Prinzipien umgesetzt wurden (situiertes Lernen, experimentierendes Lernen und sozial-konstruktivistisches Lernen (Lernen durch Handeln). Es wurden verschiedene Fragebögen eingesetzt (z. B. zur Selbstwirksamkeit) und halbstrukturierte Interviews durchgeführt. Insgesamt zeigt die Untersuchung, dass River City Grundlagenwissen und gleichzeitig Forschungsfähigkeit besser vermitteln kann als konservative (zumeist papierbasierte) Methoden. Zudem hat der Einsatz einen positiven Effekt auf das Engagement im Lernprozess und die Aufmerksamkeit (Ketelhut, Nelson, Clarke & Dede, 2010, S. 66).

Anhand der Studie wird deutlich, dass die Modifizierungen zunehmend detaillierter und weniger umfangreich wurden. Des Weiteren ist die Studie durch Merkmale von DBR charakterisiert, z. B. Einbezug von Praktikern und Praktikerrinnen (Lehrenden und Lernenden) von Beginn an (interventionsorientiert und kollaborativ), Einsatz unterschiedlicher qualitativer und quantitativer wissenschaftlicher Methoden (integrativ) sowie schrittweise formative Evaluation (iterativ). Theoretisch konnten aus der Studie Hinweise über das Engagement von Lernenden abgeleitet werden, praktisch wurde die virtuelle Lern- und Experimentierumgebung entwickelt.

Ein Praxisbeispiel aus dem deutschen Sprachraum stammt von Wilhelm u.a. (2012). Dort wurde ein Unterrichtskonzept inklusive Experimenten, Medien und Leistungstest entwickelt und evaluiert. Thematisch geht es dabei um die Vermittlung eines grundlegenden Verständnisses von Mechanik, insbesondere Geschwindigkeit, bei Grundschulkindern und ist somit in der Physikdidaktik verortet. Das Konzept unterscheidet sich von traditionellen Konzepten dadurch, dass der Fokus auf die zweidimensionale Bewegung gerichtet ist und spezielle Materialien (z. B. eine Software zur Simulation) eingesetzt werden. Es wurden mehrere Zyklen durchlaufen, wobei als Zyklen auch bereits durchgeführte Studien als Vorstudien in das Projekt einbezogen wurden und eine Hauptstudie mit drei Zyklen umgesetzt wurde. Die Zyklen der Hauptstudie führten zur kontinuierlichen Verbesserung der Materialien und zudem zu einer Anpassung an den siebten Jahrgang. Quantitative Methoden kamen in Form von Befragungen zum fachlichen Verständnis, zum Interesse am Physikunterricht sowie zu Selbstwirksamkeitserwartungen zum Einsatz. Qualitativ wurden Interviews mit Lehreren-

den sowie Lernenden und Videoanalysen durchgeführt. Des Weiteren wurden Unterrichtstagebücher von Lehrenden geführt und von den Wissenschaftlern und Wissenschaftlerinnen ausgewertet. Als praktischer Output wurde neben dem Unterrichtskonzept ein Lehrtext entwickelt. Als Ergebnis auf der theoretischen Ebene wurde festgestellt, dass der Einsatz der Materialien und die Durchführung des entwickelten Konzepts (mit dem Fokus auf zweidimensionale Bewegungen) zu einem fachlich größeren Verständnis führt als bei Personen, die mit traditionellen Methoden und Konzepten unterrichtet wurden (Wilhelm u. a., 2012).

Gemäß DBR-Prinzipien wurde in dieser Studie ein praxisrelevantes Problem identifiziert und ein theoretischer sowie praktischer Output mit mehreren aufeinander folgende Zyklen erarbeitet. Anhand der Ergebnisse wurden Konzept und Materialien evaluiert und weiterentwickelt.

Die Beispiele machen insbesondere deutlich, wie die Verortung in der Praxis sowie die Zusammenarbeit zwischen Wissenschaftlern/Wissenschaftlerinnen und Praktikern/Praktikerinnen (hier meist im schulischen Kontext) gelingen kann. Allerdings wird durch die oben genannte Analyse von Anderson und Shattuck (2012) deutlich, dass der DBR-Ansatz im deutschen Sprachraum noch wenig verbreitet ist. Hier wäre ein vermehrter Einsatz wünschenswert, um sich einerseits im Sinne von Best Practice kontextspezifisch (ausbildungssystemspezifisch) austauschen zu können aber auch Ideen zu entwickeln und zu erproben, wie der Ansatz weiter verbessert werden und sich etablieren könnte.

3.1.6 Kritik am DBR-Ansatz

„Good research demands “skepticism, commitment and detachment” (Norris, 1997), but DBR also requires comradeship, enthusiasm, and a willingness to actively support the intervention. Thus, a certain wisdom is needed to walk this narrow line between objectivity and bias. The personal skill to hold all of these attitudes simultaneously is a challenge and a defining feature of quality DBR“ (T. Anderson & Shattuck, 2012, S. 18).

Aus den vorherigen Kapiteln gehen bereits einige Kritikpunkte an dem DBR-Ansatz hervor. Im Folgenden werden wesentliche Kritikpunkte herausgestellt.

Generalisierbarkeit: Aus der Perspektive der „klassischen“ quantitativen empirischen Bildungsforschung ist ein Schwachpunkt des DBR, dass die Ergebnisse von DBR-Prozessen – wenn überhaupt – nur sehr eingeschränkt generalisierbar und replizierbar sind. Es handelt sich vorerst immer um domänen- und kontextspezifische Ergebnisse, die meistens mit einer kleinen Stichprobe gewonnen

werden. Degradierend werden sie im englischen Sprachraum teilweise auch *humble theories* genannt (Cobb u. a., 2003). So erläutert auch Walker (2006), dass das Risiko bei DBR darin besteht, aufgrund von kleinen Samples sowie der Diversität in der Praxis eher fragmentarische Ergebnisse zu erzielen als solche nach dem klassischen Wissenschaftsparadigma. Abhilfe würde eine stärkere Annäherung an klassische empirische „rigorosere“ Wissenschaft schaffen, die Ergebnisse mit hoher externer Validität liefern kann und gleichzeitig erlernbar und einsetzbar für Praktiker und Praktikerinnen ist (Walker, 2006, S. 39). Beispielsweise können die Design-Zyklen durch empirische Datenerhebungen ergänzt werden. Auch Nachfolgestudien, in denen das Design entweder in einem neuen Setting eingesetzt und evaluiert oder variiert werden (vgl. hierzu auch Gravenmeijer & Cobb, 2006) können zu einer Steigerung der Aussagekraft führen.

Involviertheit von Forschenden und Praktikern/Praktikerinnen : Ein im DBR-Diskurs bisher wenig beschriebenes Problem ist die Frage danach, wie Wissenschaftler und Wissenschaftlerinnen zusammen mit Praktikern und Praktikerinnen Forschung betreiben. Wohl aber ist diese Herausforderung aus der qualitativen Forschung bekannt.

„If a researcher is intimately involved in the conceptualization, design, development, implementation, and re-researching of pedagogical approach, then ensuring that researchers can make credible and trustworthy assertions is a challenge“ (Barab & Squire, 2004, S. 10).

Wie Barab und Squire (2004) anmerken, ist es eine große Herausforderungen sicherzustellen, dass Forschende, die in jegliche Prozesse und insbesondere in der qualitativen Datenerhebung involviert sind, zuverlässige und glaubhafte Aussagen treffen. Der mögliche Einfluss, den Wissenschaftler/Wissenschaftlerinnen auf Praktiker/Praktikerinnen und die Praxis – also auch das Untersuchungssetting – ausüben, wurde auch insbesondere in der Aktionsforschung kritisiert (vgl. Kap. 3.1.4.2). Um die Diversität und Variabilität der Praxis im Forschungsprozess berücksichtigen zu können, empfehlen Middleton et al. eine „Standardisierung der Praxis“, indem Forschungsbeteiligte beobachtet und ggf. mit Video aufgezeichnet werden (Middleton et al. 2008, 29). Das ermöglicht einerseits, den Einfluss der Forschenden auf die Praxis zu reflektieren und bei der Interpretation der Ergebnisse zu berücksichtigen. Andererseits ermöglicht die Analyse des Videomaterials ein tieferes Verständnis für den Entwicklungsprozess und die Einflussnahme der Forschenden auf den Forschungsprozess (Middleton u. a., 2008, S. 29). Nicht zuletzt bieten die Gütekriterien

qualitativer Sozialforschung eine solide Basis für die Bewertung der Qualität (vgl. Steinke, 2005 und Kap. 8.2.5).

Kausalattribution: Ein weiterer Kritikpunkt betrifft die schwer zu realisierende Kausalattribution: Das Vorgehen im Sinne des DBR-Ansatzes kann zwar Hinweise über Effekte von Interventionen (ggf. frei von systematischen Fehlern) liefern, nicht jedoch weiterführende Informationen über Faktoren, welche diesen Effekt verursachen. Somit besteht immer das Problem, dass eine Kausalattribution nicht oder nur schwer möglich ist. Um diesen Problem zu begegnen, kann auch hier der Einsatz weiterer Methoden, wie beispielsweise Interviews und Befragungen, eine Lösung bieten.

Narrativer Charakter: Ein Kritikpunkt aber auch eine Stärke ist der narrative Charakter der Studien. Zum einen gibt es keine Garantie für die allgemeingültige Wahrheit (wie im Übrigen in anderen Forschungsdisziplinen auch – nur in einem statistisch geringerem Ausmaß). Wenn ein Design als praxistauglich bewertet wird, beruht diese Bewertung nicht darauf, dass alle möglichen Alternativen getestet wurden. Eine Berücksichtigung aller Alternativen unter dem Anspruch, die Studie in realen Settings durchzuführen, ist nicht realisierbar (Shavelson, Phillips, Towne & Feuer, 2003, S. 27). Doch gerade die Idee, die Realität nicht auf ein kontrollierbares Setting zu reduzieren, birgt das Potenzial neuer Entwicklungen.

Zum anderen ist die aus dem narrativen Charakter der Studien folgende unerlässliche systematische Dokumentation ein potenzielles Risiko für die Tätigkeit zuverlässiger Aussagen. Sie ist einerseits die Grundlage der empirischen Daten. Andererseits stellt sie die intersubjektive Nachvollziehbarkeit und Transparenz der Studie sicher, um sie auch für außenstehende Personen kritisierbar zu machen. Die systematische Dokumentation folgt zwar bestimmten Regeln, dennoch gibt es auch hier keine Garantie, die Wahrheit komplett erfasst und dokumentiert zu haben. Das Risiko kann gesenkt und die Aussagekraft durch das Ausmaß gesteigert werden, indem die Studie einerseits kommuniziert, mit weiteren Wissenschaftlern und Wissenschaftlerinnen diskutiert sowie durch unabhängige Wissenschaftler und Wissenschaftlerinnen kritisiert wird (Gravenmeijer & Cobb, 2006; Shavelson u. a., 2003, S. 87). Andererseits stellt das Maß der Ausführlichkeit und die Genauigkeit der Durchführung ein wesentliches Qualitätskriterium dar (Taylor & Bogdan, 1984) (zitiert nach Gravenmeijer & Cobb, 2006, S. 87).

Die Kritikpunkte betreffen im Wesentlichen die fragwürdige Erfüllung der Generalisierbarkeit, der Kausalattribution sowie den narrativen Charakter. Begegnet wird diesen Kritikpunkten mit der Ergänzung von DBR durch klassische empirische Methoden, die den klassischen Gütekriterien standhalten. Sollte die Entwicklung dahingehen, dass der Einsatz empirischer Methoden im DBR

priorisiert wird, ist dies jedoch – unter Berücksichtigung der DBR-Prinzipien – ein Weg, den es zu hinterfragen gilt. DBR hat weder zum Ziel, die Alles-oder-Nichts-Frage „Funktioniert der Ansatz – ja oder nein?“ zu beantworten noch generalisierbare Aussagen durch experimentelle Studien zu liefern, deren Generalisierbarkeit durch die Einhaltung klassischer Gütekriterien gewährleistet wird. Im Vordergrund geht es um die Beantwortung der Frage „Wie funktioniert der Ansatz“ (Middleton u. a., 2008, S. 41) und um die Entwicklung praxistauglicher Innovationen:

„[authors of designexperiments] make new knowledge and not on the grounds that they make new realities. [...] the goal of a design experiment is to improve education per se (as opposed to just improve knowledge of education. [...]) Substantive design is appropriate to intervention research because it holds that research can change reality rather than just study it“ (Schwartz, Chang & Martin, 2008, S. 61).

Unter Berücksichtigung der primären Ziele von DBR ist die sinngemäße Legitimation – u. a. durch die Ergänzung klassischer empirischer Methoden – dieses Ansatzes weiterhin zu diskutieren.

4 Testumgebung

In diesem Kapitel werden die Vorbedingungen, Herausforderungen und die konzeptionelle Basis für die Entwicklung der Online-Testumgebung otu.lea erläutert. Hierfür wird vorerst der Projektkontext des Projekts lea. – Literalitätsentwicklung von Arbeitskräften – erläutert, in dessen Teilprojekt 3 „Self- und E-Assessment“ otu.lea entwickelt wurde, und das Ziel des Projektes vorgestellt. Anschließend werden technische sowie konzeptionelle Herausforderungen und die entsprechenden Lösungen dargestellt.

4.1 Projektkontext: lea. – Literalitätsentwicklung von Arbeitskräften

Schon vor der leo.-Studie existierten in der Förderlinie „Forschung und Entwicklung zur Alphabetisierung und Grundbildung Erwachsener“ des Bundesministeriums für Bildung und Forschung 25 Verbünde mit über 100 Teilvorhaben, die sich mit Alphabetisierung befassen. Eines dieser Verbünde ist das Projekt lea. – Literalitätsentwicklung von Arbeitskräften. Das Ziel war, eine erwachsenengerechte Förderdiagnostik für die Dimensionen Lesen, Schreiben, Mathematik und Sprachempfinden zu entwickeln. Das förderdiagnostische Instrumentarium des lea.-Projektes zielt darauf ab, dem pädagogischen (Alphabetisierungs-)Personal diagnostische Informationen zur Verfügung zu stellen, welche für die Zuweisung der potentiellen Alphabetisierungsteilnehmer/-teilnehmerinnen zu einer geeigneten Lerngruppe (Platzierung und Selektion) und für die Zusammenstellung von individuell geeigneten Förderprogrammen genutzt werden können. Voraussetzung für diese Zielerreichung in der Praxis sind grundlegende diagnostische Tätigkeiten. Hierfür wurden vier Kompetenzmodelle entwickelt, die den Bereich der Lower Rungs⁵⁶ ausdifferenzieren. Für jede Dimension wurden Alpha-Level und sog. Kannbeschreibungen definiert. Die Kannbeschreibungen sind qualitative Kurzbeschreibungen von Fähigkeiten, z. B. „Kann bei Anrede auf Großschreibung der Sie-Form achten.“ (Heinemann, 2011, S. 91). In jeder Aufgabe der Diagnostik wird mindestens eine Kannbeschreibung durch das Bearbeiten einer Aufgabe geprüft. Die Zuordnung zu den Level dient primär der Übersichtlichkeit, nicht der Zuordnung zu Hierarchiestufen. Für die Dimension Schreiben

⁵⁶ Lower Rungs bezeichnet den Kompetenzbereich des formalen Bildungsstandes unterhalb des Mindeststandards, den sog. „unteren Sprossen der Leiter“ (Brooks u. a., 2001; Dluzak, Heinemann & Grotlüschen, 2009; Wolf u. a., 2011, S. 123).

wurde insbesondere ein empirisch überprüftes mehrdimensionales Kompetenzmodell entwickelt, dessen Entwicklung im Folgenden näher erläutert wird⁵⁷.

Das Kompetenzmodell Schreiben besteht aus fünf Alpha-Level⁵⁸, die durch einzelne Kannbeschreibungen definiert sind und die Level in mehrere Kompetenzstufen differenzieren. Als theoretischen Bezugsrahmen dienten Schriftspracherwerbstheorien (Brügelmann, 2000; Frith, 1985; Reuter-Liehr, 2008; Spitta, 1997) sowie bekannte Vorbilder aus der Praxis der Schriftsprachdiagnostik (z. B. Orientierungsrahmen der Volkshochschulen, der *Gemeinsame Europäische Referenzrahmen* (GER) für Sprachen sowie die in England entwickelten Entry-Level). Diese Bezüge stellen im Einzelnen keinen hinreichenden Ansatz für die Entwicklung einer Förderdiagnostik dar, bieten aber als Synthese eine stabile Basis, um darauf aufbauend ein Kompetenzmodell für den deutschen Sprachraum im Bereich niedriger Schreibfähigkeiten zu entwickeln. So ist ein Kompetenzmodell mit 58 Kannbeschreibungen entstanden, die linear gestaffelt und in Level eingeteilt sind. Bezüglich der Einteilung in Level ist allerdings zu bemerken, dass die Entwickler und Entwicklerinnen bei Erwachsenen nicht von einem linearen Lernverlauf ausgegangen sind, sondern die Einteilung der einfacheren Handhabbarkeit dient (Heinemann, 2011, S. 91). Auf der Grundlage dieses Kompetenzmodells kann eine valide Diagnostik erstellt werden, die es ermöglicht, passgenaue Fördermaßnahmen abzuleiten. Doch wie kann dem förderdiagnostischen Anspruch einer Stärkenorientierung Rechnung getragen werden? Die Level beginnen bei dem sog. Theoretischen Nullpunkt, dem Lesen und Schreiben einzelner Buchstaben sowie Zahlen, und reichen bis hin zur Kommasetzung. Somit wird schon das Schreibenkönnen eines einzelnen Buchstabens als Kompetenz bewertet. Jede Person bekommt daher eine positive Rückmeldung ihrer vorhandenen Fähigkeiten (Heinemann, 2011, S. 92). Die Rückmeldung stellt sowohl die Stärken als auch den Förderbedarf dar. Die Entscheidung über anschließende Fördermaßnahmen soll im Dialog zwischen Kursteilnehmenden und Kursleitenden getroffen werden⁵⁹. Lernende können sich aktiv für oder auch gegen bestimmte Förderziele und/oder Fördermaterialien entscheiden. Bisher war das Lesen- und Schreibenlernen vorwiegend der Grundschule zugeordnet, erst der Förderschwerpunkt des BMBF trieb die Forschungs- und Entwicklungs-

⁵⁷ Die Entwicklung des Förderdiagnoseinstruments für die Dimension Schreiben wurde ausführlich von Alisha Heinemann in einem Sammelband des Projektes erläutert. Die folgende Darstellung ist vorwiegend an diese Darstellung angelehnt. Weitere Details sind in Grotflüsch, Kretschmann, Quante-Brandt & Wolf (2011) nachzulesen.

⁵⁸ Der Begriff Alpha-Level wurde in Anlehnung an die Entry-Level gewählt. Diese sind nach der ALL (Adult Literacy and Life Skills) -Erhebung entstanden.

⁵⁹ Dies entspricht zudem dem förderdiagnostischen Anspruch nach McManu (2008). Demnach ist eines von fünf Attributen der Förderdiagnostik die Identifizierung und Kommunikation der Lernziele und Kriterien.

bemühungen für erwachsenengerechte Arbeitsmaterialien in der Grundbildung maßgeblich voran. Im englischen Sprachraum sind Workplace Literacy (Arbeitsplatzbezogene Literalisierung) und Family Literacy (Familienbezogene Literalisierung) bereits bekannte Konzepte; in Deutschland stand die Etablierung und Entwicklung solcher Konzepte zu Beginn der Förderlinie noch am Anfang.

Um den Lernenden eine Identifikation mit dem Diagnose- und Arbeitsmaterial zu erleichtern, wurde das lea.-Universum erschaffen. Dabei handelt es sich um ein Storyboard mit verschiedenen Charakteren und Handlungssträngen, die dem Alltag von Personen mit niedrigen Lese- und Schreibfähigkeiten nachempfunden sind. Fast 57% der funktionalen Analphabeten und Analphabetinnen in Deutschland sind beruflich tätig (Grotlüschen, 2012, S. 141 und vgl. Kap. 2.3.3), insbesondere in niedrigschwelligen Berufen wie KFZ-Mechaniker/-Mechanikerin, Gebäudereinigende oder Koch/Köchin. Die Charaktere des lea.-Universums wurden unter Berücksichtigung gängiger Tätigkeiten von funktionalen Analphabeten und Analphabetinnen sowie Geschlecht und Herkunft entwickelt. Der Frauen- und Männeranteil ist gleich hoch, Frauen arbeiten in männertypischen Berufen und umgekehrt. Der Anteil an Personen mit Migrationshintergrund beträgt 43% und der Anteil an Personen mit Behinderungen 6% (Heinemann, 2011, S. 95). So gibt es z. B. Helga Bauer, eine 36-jährige Lagerarbeiterin, Leon Schmitt, ein 32-jähriger Fensterputzer und Josy Patel, eine 43-jährige Inderin, die in einer Wäscherei arbeitet.

Bereits in der Entwicklungsphase wurden Experten und Expertinnen (funktionale Analphabeten und Analphabetinnen und Kursleitende) hinsichtlich der Akzeptanz der entwickelten Diagnose- und Fördermaterialien befragt und in die Entwicklung einbezogen. Um nicht viel (Konzentrations-)Zeit der Teilnehmenden in Anspruch zu nehmen, wurden die Aufgaben möglichst kurz gehalten. In einer Aufgabe werden daher des Öfteren mehrere Kannbestimmungen geprüft. Jede Aufgabe ist dimensionsreich, d. h. zur Beantwortung einer Schreibaufgabe müssen keine Lesekompetenzen vorhanden sein. Die Aufgabeninstruktionen müssen entsprechend von Alphabetisierungskräften vorgelesen werden.

Die Items der Dimension Schreiben wurden vor dem Hintergrund der probabilistischen Testtheorie entwickelt und mit dem Raschmodell in ihrer Schwierigkeit bestimmt (vgl. z.B. Bond & Fox, 2001; Moosbrugger, 2008). Für die Skalierung wurden die entwickelten Schreibaufgaben in vier verschiedene Testhefte zusammengefasst, die alle durch ein Set mittelschwieriger Aufgaben miteinander verankert sind. Es standen zwei Hefte mit eher leichten Aufgaben (Alpha-Level 1-4) und zwei Hefte mit eher schwierigen Aufgaben (Alpha-Level 3-6⁶⁰) zur Verfügung. Anhand einer Selbsteinschätzung der Teilnehmenden wurde

⁶⁰ Bei der Entwicklung wurde zunächst von sechs Alpha-Level ausgegangen, nach der empirischen Überprüfungen fiel das sechste Level weg.

das Testheft ausgewählt. Letztlich bearbeiten insgesamt 181 Personen die Testhefte, anhand deren Ergebnisse die Raschskalierung der Items vorgenommen wurde. Bei der Skalierung galt es, die Items in eine schwierigkeitsbestimmende Reihenfolge zu bringen. Ziel war nicht, die Items zu normieren, um auf eine Grundgesamtheit schließen zu können (Grotlüschen & Heinemann, 2011, S. 111), sondern eine sinnvolle und praktikable Anordnung der Items hinsichtlich ihrer Schwierigkeit. Für die Skalierung auf Basis der Item-Response-Theory wurde ihr „prominentester Vertreter“, das Rasch-Modell, gewählt. Aus der Vielzahl der Antwortmuster kann wahrscheinlichkeitstheoretisch auf die Itemschwierigkeit geschlossen werden. Personenfähigkeiten können daher auch ermittelt werden, wenn unterschiedliche Personen unterschiedliche Aufgaben und eine unterschiedliche Anzahl von Aufgaben bearbeiten. Voraussetzung ist lediglich die Verankerung durch eine Mindestanzahl von Aufgaben (Grotlüschen & Heinemann, 2011, S. 112). Im Unterschied zur klassischen Testtheorie sagt die Summe der richtig beantworteten Aufgaben nicht gleichzeitig etwas über die Fähigkeiten aus. Die inhaltliche Beschreibung der Personenfähigkeit muss qualitativ – hier anhand der Kannbeschreibungen, anhand derer die Items operationalisiert sind – erfolgen (Grotlüschen & Heinemann, 2011, S. 112). Weitere Ausführungen zur Item Response Theory finden sich beispielsweise in Bond & Fox (2001).

Im Vergleich zur Dimension Schreiben wurden die Diagnosematerialien für die Bereiche Lesen, Mathematik und Sprachempfinden zwar evaluiert, nicht jedoch empirisch im Sinne einer schwierigkeitsbestimmenden Aufgabenhierarchie validiert.

Das Ergebnis dieser Entwicklungsarbeiten sind die lea.-Ordner (Grotlüschen, 2010; Quante-Brandt & Jäger, 2013) und das Online Diagnoseinstrument otu.lea (www.otulea.uni-bremen.de).

Die Entwicklung des computerbasierten Instruments, welches den Untersuchungsgegenstand dieser Arbeit darstellt, wird im folgenden Kapitel erläutert.

4.2 Entwicklung der Online Testumgebung otu.lea⁶¹

Die Entwicklung der Online-Testumgebung orientierte sich an den Vorarbeiten der Teilprojekte, in denen die Aufgaben für die vier Dimensionen entwickelt wurden. In dem Teilprojekt „E-Assessment“ wurden im Wesentlichen die in folgender Tabelle aufgeführten Schritte umgesetzt:

⁶¹ Bestandteile des folgenden Kapitels wurden bereits in Wolf, Koppel, Schwedes (2011), Koppel, Wolf (2014) und Wolf, Koppel (2014) veröffentlicht.

Beschreibung der Arbeitsschritte	Projektlaufzeit				Post-Projekt-Phase		
	2008	2009	2010	2011	2012	2013	2014
Sichtung vorhandener Online-Testumgebungen und E-Portfolio-Systeme							
Theoriegeleitete Sichtung und Erprobung verschiedener multimedialer und interaktiver Itemformate							
Konzeption der Testumgebung							
Prototypische Implementation erster Items							
Erster Testlauf der Online-Testumgebung							
Vollständige Implementation der Items							
Evaluation							

Tabelle 8: Arbeitsschritte im Teilprojekt zur Entwicklung von otu.lea

Zu Beginn der Projektlaufzeit wurden Online-Testumgebungen und elektronische Portfolio-Systeme gesichtet und hinsichtlich der Zielgruppe und ihres Nutzungszweckes kategorisiert. Auf der Grundlage dieser Recherche-Aktivität wurden multimediale und interaktive Itemformate bezüglich ihrer Passung zum Messgegenstand (Literalitätskompetenzen im Grundbildungsbereich Erwachsener) identifiziert und theoretisch sowie papierbasiert erprobt. Anschließend wurde das Konzept für die Online-Testumgebung erstellt. Auf der Grundlage der in Teilprojekt eins und zwei entwickelten papierbasierten Inhalte erfolgte ein Jahr nach Projektbeginn die prototypische Implementation, woraufhin im Jahr 2010 ein erster Testlauf mit prototypischen Items durchgeführt wurde. Im Jahr 2011 erfolgte die vollständige Implementation. Ab 2009 wurden über den gesamten Projektzeitraum laufende Evaluationen durchgeführt. Trotz des Projektendes im Jahr 2011 erfolgten in den Jahren 2012 bis 2014 weitere Evaluationen und Modifikationen.

Die Übertragung der Inhalte in eine computerbasierte Version stellte die Entwickler und Entwicklerinnen vor diverse Herausforderungen: Es mussten zielgruppen- sowie technologiespezifische Charakteristika berücksichtigt werden, die eine direkte 1:1-Übertragung der papierbasierten in die computerbasierte Version nicht möglich machten.

4.2.1 Herausforderungen für die Entwicklung

Das lea.-Förderdiagnoseinstrument wie auch weitere kompetenz- und förderdiagnostische Verfahren werden in der Alphabetisierung traditionell papierbasiert und durch eine Alphabetisierungskraft betreut durchgeführt. Dies erscheint vorerst notwendig, da insbesondere die *Anleitung* für die Testaufgaben häufig eine höhere Anforderung an die Lesekompetenz der Teilnehmenden stellt als die eigentliche Testaufgabe. Auch Aufgaben, die zum Beispiel einen gesprochenen Text als Stimulus beinhalten – um zwischen Schreib- und Lesefähigkeit zu unterscheiden – müssen von der/dem Diagnostiker/in vorgelesen werden. Die dadurch entstehenden Kosten und auch die sich durch die persönliche Interaktion ergebenden Zugangsbarrieren (funktionale Analphabeten und Analphabetinnen berichten von einem Schamgefühl, sich einer anderen Person in einer Testsituation zu stellen bzw. das Aufsuchen einer Institution wie der VHS wird als ähnlich belastend wie Besuche bei einer Behörde empfunden; (vgl. z.B. Döbert u. a., 2000; Egloff, 1997; Füssenich, 2004; Schladebach, 2007) und resultieren in einer geringen Teilnehmer-/Teilnehmerinnenquote an Maßnahmen der Alphabetisierung: Obwohl die Anzahl funktionaler Analphabeten und Analphabetinnen in Deutschland 7,5 Millionen beträgt, nehmen bundesweit aber nur ca. 30.000 funktionale Analphabeten und Analphabetinnen an Alphabetisierungskursen teil (Huntemann & Reichart, 2011, S. 30). Es ist somit wünschenswert, den Einstieg in die Teilnahme an Maßnahmen zur Alphabetisierung zu erleichtern als auch genauer zu diagnostizieren, welche Kompetenzen gefördert werden können. Für diesen Bedarf erscheint es vorteilhaft, die Kompetenz- und Förderdiagnostik in der Alphabetisierung computerbasiert zu implementieren. Ein anonymer Zugang und die Möglichkeit einer anonymen, stärkenorientierten Kompetenzdiagnostik soll Betroffene dazu ermutigen, sich mit dem Thema – ihren literalen und mathematischen Kompetenzdefiziten – zu befassen. Idealerweise werden Betroffene durch die stärkenorientierte Rückmeldung ermuntert, eine Weiterbildungsinstitution aufzusuchen oder sich einer Person anzuvertrauen, mit der weitere Schritte zur Alphabetisierung besprochen und begangen werden können.

Für die Entwicklung eines computerbasierten Diagnostikprogramms standen die Entwickler und Entwicklerinnen vor zwei Herausforderungen: Einerseits bedürfen die funktionalen Analphabeten und Analphabetinnen besonderer Unterstützung, um Instruktionen und Fragen zu decodieren und diese schließlich bearbeiten zu können. Beim Einsatz der papierbasierten Diagnostik würde dies zum Teil – je nach Aufgabe – die Alphabetisierungskraft übernehmen. Dies wäre zumindest nicht möglich, wenn das computerbasierte Diagnoseinstrument alleine genutzt werden soll. Andererseits ist bei funktionalen Analphabeten und Analphabetinnen, wie bereits in Kap. 2.3.5 dargelegt, von einer niedrigen ICT-Literacy auszugehen (Niesyto, 2009, S. 6).

Für die Entwicklung von otu.lea galt es daher, aufgrund der niedrigen Literalität und der tendenziell niedrigen ICT-Literacy auf eine leichte Erschließung der Aufgabe, den Abbau von Hemmschwellen und die Förderung der Motivation zu achten. Eine reine 1:1-Übertragung von einer papierbasierten in eine digitale Form wird teilweise als wenig problematisch angesehen (Mason, Patry & Bernstein, 2001) kann aber einen Einfluss auf die Inhaltsvalidität des Instruments haben. Es gibt Hinweise darauf, dass Effekte bei der Übertragung auftreten und diese insbesondere stark bei Personen mit niedriger Literalität ausfallen (Pomplun & Custer, 2005) (vgl. Kap. 2.5 zur Äquivalenzproblematik). Die größte Herausforderung für die Entwickler und Entwicklerinnen bestand jedoch darin, dass die betreuende Alphabetisierungskraft im Falle einer anonymen Nutzung der computerbasierten Diagnostik wegfällt; mündliche Erläuterungen und die Instruktion der Aufgaben sind nicht möglich. Dies kann dahingehend einen Einfluss auf die Inhaltsvalidität haben, da es für eine Person mit sehr niedriger Literalität ggf. schwieriger ist, die Anleitung zu lesen als die Aufgabe zu bearbeiten, sie die Aufgabe aber ohne Anleitung nicht lösen könnte. Kompensatorisch kann hier eine multimediale Ergänzung durch die auditiv dargebotene Erläuterung der Aufgabe wirken. Zum einen ist für das Verständnis der Instruktion keine Lesekompetenz notwendig. Zum anderen erfolgt die Informationsweitergabe über zwei Kanäle (auditiv und visuell) und kann daher das Verstehen des Aufgabentextes erleichtern (vgl. Kap. 2.4 zur CLT und CTML). Ziel war es somit im Sinne einer sog. *Accomodation*, die Bearbeitung einer Aufgabe medial zu unterstützen, ohne das zu messende Konstrukt zu verändern oder zu vereinfachen (Laitusis, 2010, S. 153)⁶². Demnach galt es, kontextsensitive Hilfefunktionen⁶³ und Hinweissignale (z. B. blinkende Schaltflächen – Cueing) einzubinden. Die technologische Umsetzung und das Konzept der Online-Testumgebung werden in den folgenden Kapiteln erläutert. Die visuelle Umsetzung von otu.lea ist Bestandteil des empirischen Teils und wird entsprechend dort erläutert.

4.2.2 Das Konzept von otu.lea

Aufgrund der tendenziell niedrigen ICT-Literacy ist die Benutzer-/Benutzerinnenführung von besonderer Bedeutung. Eine angemessene Nutzer-/Nutzerinnenführung macht ein computerbasiertes Diagnoseinstrument im Ideal-

⁶² Wird durch die Veränderung der Aufgabe auch das Konstrukt verändert, handelt es sich um *modification*.

⁶³ „Kontextsensitive Hilfefunktion“ bedeutet, der Hilfetext oder der Umfang der Hilfethemen hängt von der Arbeitsaufgabe der Nutzer und Nutzerinnen, der letzten Eingabe, dem ausgewählten Objekt, der gegenwärtigen Position oder dem aktuellen Modus des Systems oder der Anwendung ab (DIN 2011, 42).

fall leicht bedienbar und damit besonders zugänglich für die Zielgruppe. Otu.lea besteht aus mehreren Teilbereichen, die nacheinander durchlaufen werden können:

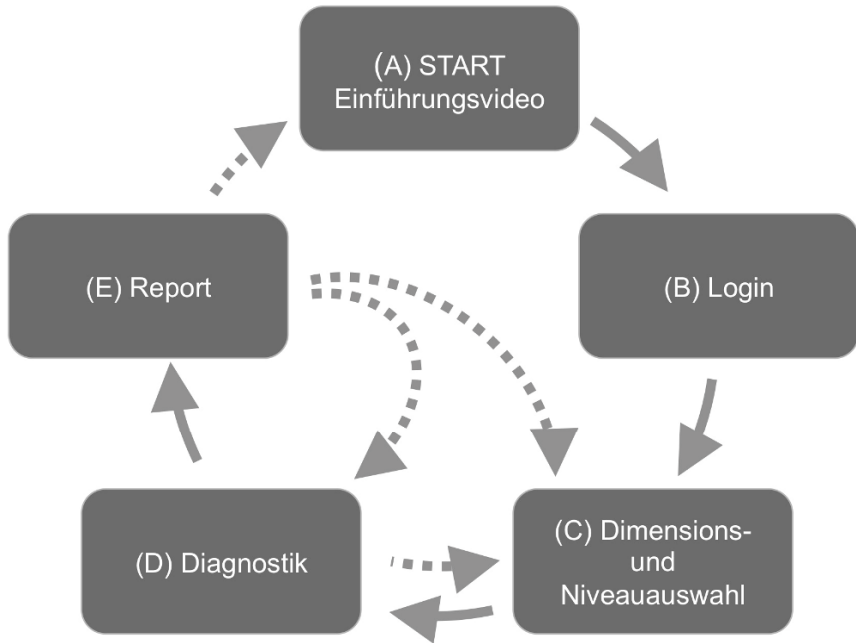


Abbildung 12: Ablauf otu.lea

(A) In einem kurzen Einführungsvideo werden otu.lea, der Aufbau, die Symbole und grundlegende Funktionen erläutert. (B) Nachdem die Personen sich mit einem anonymisierten Code angemeldet haben, wählen sie (C) die Dimension (Lesen, Schreiben, Mathematik oder Sprachempfinden) aus und nehmen eine Selbsteinschätzung für die Niveaueauswahl vor (es stehen die „leicht“, „mittel“ und „schwer“ zur Verfügung, die auditiv unterstützt mit Beispielen versehen sind). Anschließend wird die Diagnostik durchlaufen (D). Während der Diagnostik besteht die Möglichkeit, wieder zur Dimensions- und Niveaueauswahl zurückzukehren (z. B. falls Teilnehmende feststellen, sich nicht angemessen eingeschätzt zu haben). (E) Nach der Diagnostik erhalten die Teilnehmenden eine individuelle, stärkenorientierte Rückmeldung in leichter Sprache, die auf dem otu.lea-Server an der Universität Bremen (<http://otulea.uni-bremen.de>) dynamisch für jeden

Testdurchgang erstellt wird. Die Rückmeldung wird zum anderen direkt in der Testumgebung auditiv angeboten, kann also auch wieder vorgelesen werden. Darüber hinaus haben die Teilnehmenden die Möglichkeit, den Report auf dem Computer als PDF-Dokument zu speichern. Die selbstständigen Nutzung und der Möglichkeit, ihre Ergebnisse und somit ihren Lernverlauf selbstständig zu evaluieren hat idealerweise einen motivationsfördernden Effekt (Ally, 2004; S. K. Reed, 2006; Svinicki, 1999).

Über einen separaten Zugang auf der Homepage von otu.lea können Kursleiter und Kursleiterinnen einen detaillierten Report mit einer differenzierten Auflistung der Kompetenzen abrufen, von denen die Ableitung individueller Fördermaßnahmen möglich sein soll. Teilnehmende können sich beliebig oft mit dem gleichen Code anmelden. Haben sie mehrfach Aufgaben bearbeitet, wird im Kursleitendenreport die Entwicklung im Vergleich zum vorherigen Durchlauf abgebildet. Somit haben Kursleitende u. a. die Möglichkeit, den Effekt von Fördermaßnahmen zu prüfen.

PHASE III – Konzeptevaluation

5 Das Forschungsdesign

Als Grundlage für das Forschungsdesign dient der Design-Based Research-Ansatz, der durch systematische Gestaltung, Durchführung, Evaluation und Re-Design die Entwicklung von innovativen Lösungen für Probleme in der Bildungspraxis verfolgt (vgl. Kap.3). In den folgenden Kapiteln wird die Forschungsfrage konkretisiert, mit den Grundlagen und Charakteristiken des DBR-Ansatzes verknüpft und vor dem Hintergrund des DBR-Ansatzes dargestellt um damit das Forschungsvorhaben zu legitimieren und anschließend dessen Umsetzung darzustellen.

5.1 Entwicklung der Forschungsfrage und des Forschungsdesigns

Das Vorgehen wird sowohl Ex-Ante (Abbildung 13 und Abbildung 14) als auch Ex-Post (in Kap. 5.3 anhand der Abbildung 16 und der Tabelle 9) abgebildet, um den Prozess der Forschungsdesignentwicklung intersubjektiv nachvollziehbar darzustellen.

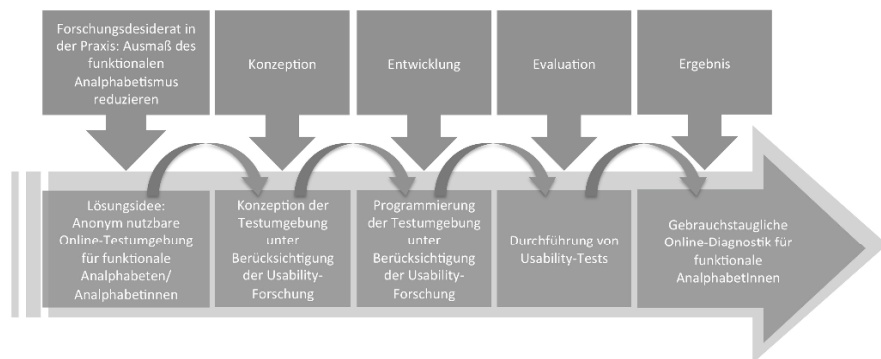


Abbildung 13: Forschungsdesign Ex-Ante - Version 1

Das Forschungsvorhaben wurde im Jahr 2009 begonnen. Der Anlass für das Forschungsvorhaben ergab sich aus dem Problem des Funktionalen Analphabetismus in Deutschland sowie den Herausforderungen in der Alphabetisierungspraxis. Einen Beitrag zur Reduzierung dieses Problems sollte ein computerbasiertes Diagnoseinstrument leisten. Das Entwicklungsziel bestand in der Programmierung einer validen Online-Diagnostik für funktionale Analphabeten und Analphabetinnen. Der vorläufige „grobe“ Ablauf sah somit die Phasen der Kon-

zeption, der Entwicklung, der Evaluation sowie der Implementation vor. Die Gebrauchstauglichkeit besaß bei der Entwicklung eine besondere Relevanz, da die Online-Testumgebung dem Anspruch der Validität genügen musste und es zu vermeiden galt, dass die Computerkompetenz einen Einfluss auf die Performanz hat. Allerdings ist die theoretischen Grundlage bezüglich der Usability-Anforderungen im Grundbildungsbereich sehr gering, so dass in der Berücksichtigung zielgruppenspezifischer Usability-Anforderungen eine Herausforderung bestand. Das Entwicklungsziel traf somit auf „Neuland“. Daraus wurde ein erstes (neben dem Entwicklungsziel) zusätzliches Kriterium für das Forschungsdesign abgeleitet: Um sich an eine Lösung heranzutasten musste iterativ mit mehreren Entwicklungs- und Evaluationsschleifen vorgegangen werden:

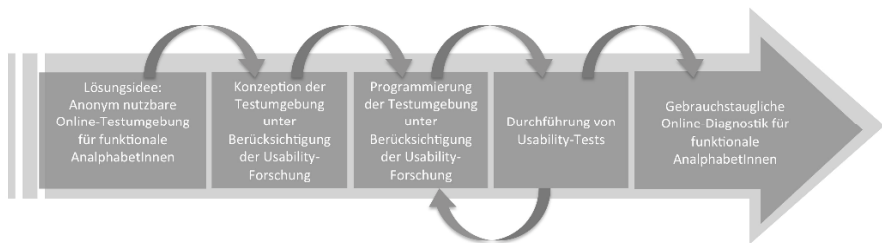


Abbildung 14: Forschungsablauf Ex-Ante - Version 2

Das Vorgehen entspricht dem typischen agilen Software-Entwicklungsprozess (vgl. z. B. Braude & Bernstein, 2011) und dem Kriterium der Usability-Entwicklung „*Test early and often*“ (vgl. Kap. 6).

Aufgrund mangelnder Erfahrung mit der Zielgruppe bestand des Weiteren das Risiko, an den potenziellen Nutzern und Nutzerinnen und dem Bedarf „vorbei“ zu entwickeln.

Entsprechend sollte die erste Evaluationsschleife mit einer Methode der partizipativen Design-Entwicklung durchgeführt werden. Aufgrund des frühen - vornehmlich noch konzeptionellen - Entwicklungsstadiums wurde Paper Prototyping-Experiment als adäquate Methode identifiziert. Im Anschluss sollten weitere Heuristische Evaluationen und Usability-Tests realisiert werden. Im Kontext des Paper Prototyping-Experiment wurden erste Erfahrungen mit den potenziellen Nutzern und Nutzerinnen gesammelt⁶⁴. Es stellte sich heraus, dass die potenzielle Nutzer-/Nutzerinnengruppe und dessen Eigenschaften in dem bisherigen Design zu wenig berücksichtigt wurden. Eine Qualitätssteigerung und die Nützlichkeit des Konzepts konnte nur mit potenziellen Nutzern und Nutzerinnen zusammen entwickelt werden. An das Forschungsdesign stellte sich somit neben dem Kriterium des iterativen Vorgangs die Anforderung, Personen aus der

⁶⁴ Das Paper Prototyping-Experiment wird in Kap. 6 eingehend erläutert.

Praxis stärker einzubeziehen. Aus den eben explizierten Herausforderungen (wenig Erkenntnisse aus der Forschung, keine bestehenden zielgruppenspezifischen Usability-Anforderungen, wenig Erfahrungen mit der Zielgruppe) konnte nicht davon ausgegangen werden, das Entwicklungsziel ohne inkrementelle und partizipative Anteile zu erreichen. Ausschließlich experimentelle Designs würden aufgrund der notwendigen Reduktion der Untersuchungsbedingungen den komplexen Anforderungen (z. B. heterogene Personengruppen und Computerkenntnisse) nicht gerecht werden. Auch die Evaluationsforschung würde dem Forschungsziel nicht gerecht, da diese nicht das Ziel verfolgt, neue Theorien zu entwickeln (vgl. Kap. 0). Es musste ein Forschungsansatz gewählt werden, der die Fokussierung sowohl eines praktischen als auch theoretischen Output integriert und die iterative Entwicklung ermöglicht. Der DBR-Ansatz lieferte mit seinen Merkmalen und Kriterien der iterativen Vorgehensweise, der Einbeziehung der potenziellen Nutzer und Nutzerinnen und der Generierung eines neuen theoretischen Verständnisses einen adäquaten Rahmen für das Forschungsdesign. Im Folgenden wird DBR im Kontext der Fragestellung näher erläutert.

5.2 Ziele des Forschungsvorhabens

Zentrale Ziele dieser Arbeit sind zielgruppenspezifische Anforderungen zu explorieren, Möglichkeiten und Herausforderungen computerbasierter Kompetenzdiagnostik im Grundbildungsbereich zu definieren sowie Heuristiken für die Usability einer Online-Testumgebung im Grundbildungsbereich zu entwickeln.

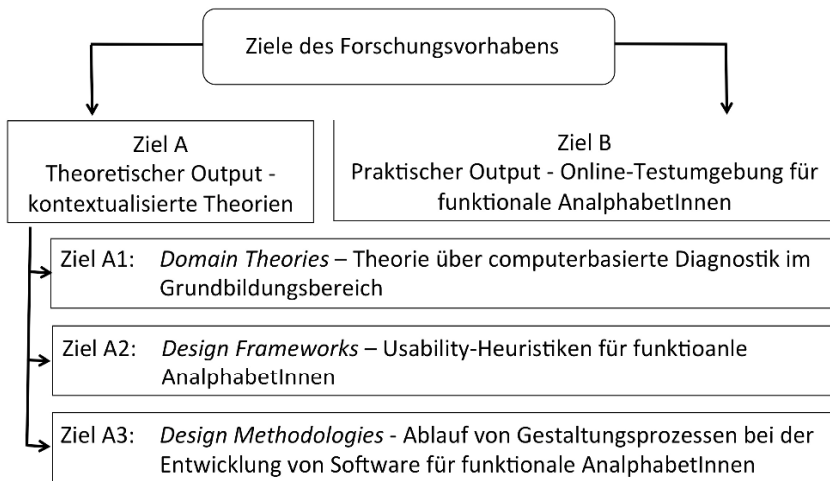


Abbildung 15: Ziele des Forschungsvorhabens in Anlehnung an den DBR-Prozess

Mit dem DBR-Ansatz und der Umsetzung von Zyklen, bestehend aus systematischer Gestaltung, Analyse, Überprüfung und das Re-Design, wird sowohl ein theoretischer (Ziel A) als auch ein praktischer Output fokussiert (Ziel B).

Die übergeordneten Ziele sind die Entwicklung kontextualisierter bzw. bereichsspezifischer Theorien (Ziel A) und die Entwicklung der Online Testumgebung für funktionale Analphabeten und Analphabetinnen otu.lea (Ziel B). Ziel A ist in 3 Teilziele auszdifferenzieren:

Ziel A1: Hinsichtlich der *Domain Theories* wird das Ziel verfolgt, (eine) bereichsspezifische Theorie(n) über computerbasierte Diagnostik im Grundbildungsbereich zu entwickeln. Einerseits werden die Herausforderungen und Möglichkeiten, die durch den Einsatz von computerbasierter Diagnostik im Kontext förderdiagnostischen Handelns im Grundbildungsbereich entstehen, reflektiert und anschließend dem Versuch einer Generalisierung unterzogen. Andererseits sollen die spezifischen Herausforderungen, Funktionalitäten sowie die multimediale Anreicherung, die das Interface von otu.lea bietet, reflektiert und theoretisiert werden. Konkretes Ziel ist die Entwicklung einer Theorie über valide computerbasierte Diagnostik im Grundbildungsbereich. Die Datengrundlage bildet hierfür das gesamte Material aus Videoaufzeichnungen, Interviews, Befragungen sowie des Data-Logging.

Ziel A2: Das *Design Framework* soll aus einem Katalog von Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen bestehen. Diese beinhalten die zielgruppenspezifischen Design-Anforderungen für die Online-Testumgebung. Die Anforderungen beziehen sich insbesondere auf die Interface- und Itemgestaltung sowie auf die Gestaltung von Hilfsfunktionen und Animationen. Als Datengrundlage werden hierfür primär die Interviews herangezogen.

Ziel A3: Die Erstellung einer *Design Methodology* wird in dieser Arbeit nur randständig behandelt. Es wird geprüft, ob eine Methodologie formuliert werden kann, welche den generalisierten Ablauf der Entwicklung eines computerbasierten Diagnoseinstruments beschreibt. Dafür wird die systematische Beobachtung und Dokumentation der prozessorientierten Entwicklung analysiert.

Wie bereits angemerkt, ist die Zielerreichung von den Ergebnissen der Evaluationszyklen abhängig. Ob die Entwicklung von bereichsspezifischen Theorien, eines Design Frameworks und einer Design Methodology realisierbar sind, wird im Anschluss an die Ergebnisauswertung kritisch hinterfragt.

Dem DBR-Ansatz folgend werden für die Zielerreichung die Aufgaben (Items von otu.lea), die Ziele (valide Diagnostik), Prozesse (Lösungsprozess bzw. Teilnahmeprozess) und die Akteure sowie Akteurinnen (Funktionale Analphabeten und Analphabetinnen und deren Anforderungen an ein computerbasiertes Diagnoseinstrument) einbezogen. Zusammenfassend stellt die Usability (und als Teil dessen das Design) von otu.lea den zentralen Untersuchungsgegenstand dar.

5.3 Ablauf des dritten Zyklus` und Forschungsdesign

Der Ablauf und das Vorgehen orientierten sich an dem DBR-Ansatz. Die Phasen und Zyklen werden durch Methoden der empirisch-pädagogischen Forschung ergänzt. Das Phasen-Modell nach Reeves bildet das Grundgerüst.

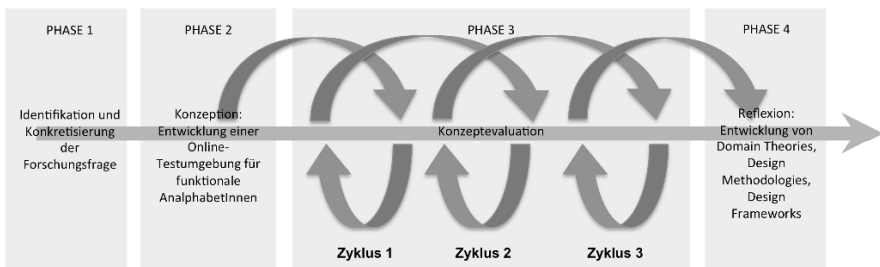


Abbildung 16: Schematischer Ablauf des DBR-Forschungsprozesses in Anlehnung an Reeves (2006) Ex-Post

Der Forschungsprozess zeichnet sich durch eine iterative Vorgehensweise mit vier Phasen aus: In der ersten Phase wird die Forschungsfrage identifiziert und konkretisiert. Die zweite Phase dient der Konzeption – der Entwicklung eines Prototyps der Online-Testumgebung. In der dritten Phase werden drei Design-Zyklen durchlaufen, bestehend aus Gestaltung, Durchführung, Analyse und Re-Design. Schließlich wird in der vierten Phase der Forschungsprozess reflektiert, die Ergebnisse analysiert und hinsichtlich der Zielsetzung dieser Arbeit aufbereitet. Die inhaltliche Konkretisierung des Forschungsdesigns zeigt die folgende Abbildung.

P 1	P 2	P 3			P 4	
Identifikation und Konkretisierung der Forschungsfrage	Konzeption		Zyklus 1: Paper Prototyping	Zyklus 2: Usability-Studie 1	Zyklus 3: Usability-Studie 2	Ziele 1. Praktischer Output 2. Kontextualisierte Theorien
		Gestaltung	Theoretische Vorüberlegungen (u.a. Definition der experimentellen Bedingungen), Auswahl und Akquirierung der Zielgruppe	Theoretische Vorüberlegungen, Auswahl und Akquirierung der Zielgruppe	Theoretische Vorüberlegungen (u.a. Definition der experimentellen Bedingungen), Auswahl und Akquirierung der Zielgruppe	
			Auswahl der zu testenden Items, Entwicklung der papierbasierten Items	Auswahl des Itemsets	Auswahl des Itemsets	
		Durchführung der Usability-Studien	Institution: VHS mit Alphabetisierungskurs Methoden: ☐ Paper Prototyping-Experiment ☐ Befragung	Institution: VHS mit Alphabetisierungskurs Methoden: ☐ Usability-Testing (Video und Data-Log) ☐ Interviews	Institutionen: 4 Einrichtungen mit Alphabetisierungskursen Methoden: ☐ Usability-Testing (Video und Data-Log) ☐ Interviews ☐ Fragebögen zur User Experience und Computerkompetenz	
			Vorstellung, Erläuterung des Projektkontextes und der Spielregeln	Vorstellung, Erläuterung des Projektkontextes, des Ablaufs und der Bedingungen	Vorstellung, Erläuterung des Projektkontextes, des Ablaufs und der Bedingungen	
		Analyse	Auswertung der Videoaufzeichnung	Auswertung der Performance, der Interviews, der Videos und Beobachtungen	Auswertung der Performance, der Interviews, der Videos und Beobachtungen	
			Definition der zu überarbeitenden Aspekte	Definition der zu überarbeitenden Aspekte	Definition der zu überarbeitenden Aspekte	
		Re-Design	Modifikation von otu.lea	Modifikation von otu.lea	Abschließende Überarbeitungen	

Tabelle 9: Konkretisierte Darstellung des Forschungsdesigns Ex-Post

Phase 1 (P1) – Identifikation und Konkretisierung der Forschungsfrage: In der ersten Phase wird die Forschungsfrage konkretisiert. Rückblickend wurde diese Phase mit der theoretischen Herleitung (Kap. 2) dieser Arbeit umgesetzt. Anlässlich des Problems des Funktionalen Analphabetismus (Kap. 2.3.2 und 2.3.3) und der Herausforderungen in der Alphabetisierung (Kap. 2.3.4) lautet die Forschungsfrage: Wie ist eine Online-Testumgebung zu konzipieren, um sie für

die Zielgruppe gebrauchstauglich und zudem valide zu gestalten? Eine genaue Erläuterung der Ziele findet sich im vorherigen Kap. 5.1.

Phase 2 (P2) – Konzeption: In der zweiten Phase wird das Konzept für die Online-Testumgebung entwickeln. Der Entwicklungsstand zu Beginn des Forschungsvorhabens wurde in Kap. 4 erläutert.

Phase 3 (P3) – Konzeptevaluation: In der dritten Phase wird das Konzept evaluiert und modifiziert. Im Sinne des DBR-Ansatzes besteht die Entwicklung und Überprüfung von otu.lea aus mehreren – nämlich drei – Design-Zyklen. Die Zyklen 1 und 2 stellen die Vorstudien dar und werden daher verkürzt in ihren einzelnen Zyklusphasen und Ergebnissen aufgeführt. Der dritte Zyklus ist die Hauptuntersuchung.

Zyklus 1 (2010) – Paper Prototyping: Im Zyklus 1 wurde ein erster Papierprototyp mit einem Paper Prototyping-Experiment evaluiert. Der erste Zyklus ist die erste Vorstudie, deren Vorgehen und Ergebnisse in Kap. 6 komprimiert vorgestellt werden.

Zyklus 2 (2012) – Usability-Studie 1: Auf der Grundlage des ersten Design-Zyklus wurde das Konzept von otu.lea modifiziert und es wurde ein technischer Prototyp entwickelt. In diesem zweiten Design-Zyklus – der zweiten Vorstudie – wurde der technische Prototyp überprüft. In Kap. 7 werden das Vorgehen, die Methoden und die Ergebnisse vorgestellt. Die Phase des Re-Designs im zweiten Zyklus diente der Weiterentwicklung der Online-Testumgebung.

Zyklus 3 (2013) – Usability-Studie 2: Der dritte Design-Zyklus stellt die Hauptuntersuchung dar. In diesem Zyklus wurde otu.lea einer ausführlichen Evaluation unterzogen. Im 8. Kap. werden das Vorgehen, die Methoden, der Untersuchungskontext, der Auswertungsmethoden und die Ergebnisse dargestellt, analysiert und interpretiert.

Die Zyklen bestehen aus jeweils vier Phasen (vgl. Kap. DBR Grundlagen). Für die Übertragung dieses Ansatzes auf die Forschungsfrage sind die Phasen mit den Schritten der Usability-Evaluation in Verbindung zu setzen.

Zyklen von DBR	Schritte der Usability-Evaluation (in Anlehnung an Niegemann 2008)
a) Gestaltung	Detaillierte Planung; Zusammenstellung des Testmaterials
b) Durchführung	Durchführung der Usability-Tests; Instruktion der Teilnehmenden
c) Analyse	Datensammlung und Analyse; Berichtlegung und Zusammenstellung
d) Re-Design	Re-Design: Umsetzung der Evaluationsergebnisse ⁶⁵

Tabelle 10: DBR-Zyklen und Schritte der Usability-Evaluation

- a) Gestaltung: Zur Gestaltungsphase gehört im Kontext der Fragestellung und in Verbindung der Anforderung an einer Usability-Evaluation die Konzeption eines theoretischen Rahmens. Die Gestaltungs-Phase beinhalten die detaillierte Planung der Usability-Studie sowie die Zusammenstellung des Testmaterials. Hierfür wurden auf der theoretischen Grundlage die Kriterien entwickelt, welche die Evaluationsgrundlage bilden. Den Evaluationsgegenstand stellt otu.lea und insbesondere das Interface in dem jeweiligen Entwicklungsstadium dar. In der ersten Zyklusphase wurde zudem das Test-Design für den jeweiligen Zyklus entwickelt. Hierzu zählen die Definition der Zielgruppe, die Auswahl der zu überprüfenden Heuristiken, die Gestaltung und Auswahl der Items sowie die Auswahl der Methode und -technik.
- b) Durchführung: In der Durchführungsphase wurde die Usability-Studie umgesetzt. Das entwickelte Design wurde in der Praxis erprobt. Die Interaktion zwischen Medien (otu.lea) und Probanden/Probandinnen (funktionalen Analphabeten und Analphabetinnen) sowie zwischen Probanden/Probandinnen und Forschenden wurde durch Data-Logging, Beobachtungen und Videoaufzeichnungen erfasst, so dass systematische Effekte beobachtbar sind.
- c) Analyse: Die Phase der Analyse dient der Datensammlung und -analyse sowie der Berichtlegung und Zusammenstellung der zu überarbeitenden Aspekte. Die Analyse erfolgte anhand der Kriterien des theoretischen Rahmens und auf der Datengrundlage. Das Design wurde überprüft und es wurden Modifizierungsmöglichkeiten abgeleitet.
- d) Re-Design: In der Phase des Re-Designs wurde der Evaluationsgegenstand modifiziert. Dies geschah auf der Grundlage der gewonnenen Beobachtungen und den daraus abgeleiteten Modifizierungsvorschlägen.

⁶⁵ Die Phase des Re-Designs stellt keinen Bestandteil der Evaluations-Schritte nach Niegemann (2008) dar.

In der Methodenwahl ist eine Heuristische Evaluation nicht berücksichtigt. Das mag vordergründig irritierend erscheinen, da es in dem Forschungsvorhaben u. a. um die Überprüfung von Heuristiken geht. Begründet ist der Ausschluss von heuristischen Evaluationen aus diesem Forschungsvorhaben, da es sich bei der Methode um ein informelles Verfahren handelt, bei dem die potenziellen Nutzer und Nutzerinnen nicht einbezogen werden (Niegemann, 2008, S. 434). Auch geht es in dieser Arbeit nicht nur um die Überprüfung der Heuristiken, sondern auch um die mögliche Notwendigkeit, weitere Heuristiken, die nicht Bestandteil der Evaluation sind, formuliert werden. Dies würde in einer heuristischen Evaluation nicht berücksichtigt werden.

Phase 4 (P4) – Reflexion: In der vierten Phase ist einerseits das Ergebnis des letzten (dritten) Design-Zyklus zu verorten: Die Online-Testumgebung. Andererseits sollen auf der Grundlage der Ergebnisse kontextualisierte Theorien entwickelt werden. Fokussiert wurde die Formulierung von Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen. Dies wurde in den Kap. 2.6.5 und wird in Kap. 8.6 behandelt. Es galt zu überprüfen, ob die Design-Prinzipien (die Heuristiken) auf weitere Kontexte (z. B. auf weitere Zielgruppen) übertragen werden können. Auch war zu überlegen, ob das generierte Wissen für weitere theoretische Erkenntnisse genutzt werden kann.

5.4 Forschungsstrategisches und -methodisches Vorgehen

Forschungsstrategisch und -methodisch wurde im Rahmen dieser Arbeit folgendermaßen vorgegangen (in Anlehnung an Edelson, 2002):

- a) **Forschungsbasierte und -orientierte Vorgehensweise:** Die theoretische Basis für die Beantwortung der Forschungsfragen liefern Erkenntnisse aus der Pädagogischen Diagnostik (vgl. Kap. 2.1), der Forschung über computerbasierte Diagnostik (vgl. Kap. 2.2), der Alphabetisierung (Kap. 2.3), der CLT und der CTML (vgl. Kap. 2.4) sowie der Usability-Forschung (vgl. Kap. 2.6). Zudem bilden die Ergebnisse aus dem lea.-Projekt einen Teil der theoretischen Basis und liefern aktuelle Bezüge zur Alphabetisierungsforschung.
- b) **Systematische Dokumentation:** Neben der permanenten kritischen Überprüfung der Entwicklungsschritte wurden die Ergebnisse der Evaluationen systematisch dokumentiert und aufbereitet. Aus den Vorstudien werden die wesentlichen Ergebnisse in den folgenden Kapiteln dargestellt, um intersubjektiv nachvollziehbar die Vorgehensweise in der Hauptstudie zu begründen. Zudem wurde die Entwicklung von otu.lea in mehreren Publikationen dokumentiert, reflektiert und damit öffentlich zugänglich gemacht (Koppel, Küster & Wolf, 2013; Wolf & Koppel, 2010, 2014; Wolf u. a., 2011).

- c) Formative Evaluation: Durch die Evaluationen des Paper-Prototyping-Experiments sowie der Usability-Tests 2012 wurden Unzulänglichkeiten identifiziert und in der weiteren Entwicklung berücksichtigt. Dies geschah sowohl auf der inhaltlichen Ebene (Iteminhalte und -auswertung) als auch auf der Usability-Ebene (Modifizierung des Interfaces und der Benutzer-/Benutzerinnenführung).
- d) Generalisierung: Die Generalisierung erfolgt in Kap. 8.6. Hierbei ging es darum, die systematisch dokumentierten Ergebnisse (der Problemanalysen des Lösungs- und Designprozesses) miteinander in Beziehung zu setzen und in einen übergeordneten Kontext zu übertragen (Usability-Anforderungen im Grundbildungsbereich). Wie bereits oben erwähnt wurden die Entwicklungen einer Domain Theory und eines Design Frameworks fokussiert.

Die Motivation für diese Arbeit liegt in dem Bestreben nach der Verbesserung und Optimierung der Online-Testumgebung, um einen Beitrag für die Alphabetisierung zu leisten. Durch die Entwicklung des Kompetenzmodells im lea.-Projekt (vgl. Kap. 4.1) wurde bereits ein Beitrag für die Bildungspraxis im Alphabetisierungskontext geliefert. Es galt, auch otu.lea für Alphabetisierungskräfte und Funktionale Analphabeten und Analphabetinnen zugänglich zu machen und gebrauchstauglich zu gestalten. Die Motivation für die Verwendung des DBR-Ansatzes ist in dem notwendigen iterativen Vorgehen mit mehreren Design-Zyklen sowie der Einbeziehung der potenziellen Nutzer und Nutzerinnen und die Generierung eines theoretischen Verständnisses begründet (vgl. auch Kap. Entwicklung und Entfaltung der Forschungsfrage).

Die Gütekriterien der empirischen Forschung und des DBR (Neuheit und Nützlichkeit) werden folgendermaßen erfüllt:

Neuheit: Neu ist die Umsetzung eines computerbasierten Diagnoseinstruments für den Grundbildungsbereich. Es bestehen zwar computerbasierte Lernprogramme, jedoch keine Online-Testumgebungen, die eine individuelle und differenzierte Diagnose der literalen und mathematischen Grundkompetenzen bieten. Auch in nicht deutschsprachigen Ländern sind keine derartigen Messinstrumente bekannt. Somit wird der Anspruch der Neuheit sowohl auf inhaltlicher als auch auf „technischer“ Ebene erfüllt: Inhaltlich ist die Entwicklung eines Kompetenzmodelles für den Grundbildungsbereich Erwachsener eine neue Entwicklung im deutschsprachigen Raum. Auf der technischen Ebene ist die Realisierung einer Online-Testumgebung und zudem die Anreicherung dieser Umgebung mit multimedialen Cues innovativ.

Nützlichkeit: Die Nützlichkeit besteht in einem ressourcenschonenden und Hemmschwellen senkenden (vgl. Kap. 4.2) Instrument, welches frei zugänglich zur Diagnose und Förderung genutzt werden kann. Zum anderen wird in dieser Arbeit ein wissenschaftlicher Anspruch verfolgt, indem die qualitativen und

quantitativen Daten nach methodologisch abgesicherten Verfahren ausgewertet werden.

Einhaltung der Gütekriterien empirischer Forschung: Der DBR-Ansatz wurde mit statistischen Verfahren der empirischen Sozialforschung ergänzt und in das Forschungsdesign integriert. Dem Mixed Method-Ansatz folgend wurden qualitative und quantitative Datenerhebungsinstrumente miteinander kombiniert.

„Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration“ (Johnson, Onwuegbuzie & Turner, 2007, S. 123).

Vorgegangen wurde dabei im Sinne eines *Emergent Mixed Method Design* (vgl. z. B. Creswell & Clark, 2010, S. 54), bei dem die Auswahl der Instrumente unter Berücksichtigung vorheriger Studien getroffen wurde. Es stand also nicht zu Beginn des DBR-Prozesses fest, welche Instrumente im dritten Zyklus zum Einsatz kommen. Dabei wurde entsprechend der Gütekriterien der empirischen Forschung vorgegangen. Diese werden zusammen mit den Erhebungsinstrumenten in Kap. 8.2 erläutert.

Der Fokus liegt somit auf der Entwicklung und Innovation – die Vorgehensweise bedient sich empirischer Verfahren, um otu.lea zu einem validen Erhebungsinstrument in der Bildungspraxis zu machen.

5.5 Kritik am DBR-Ansatz im Kontext der Fragestellung

Der Ansatz des DBR verleiht dem Forschungsvorhaben einen handhabbaren Rahmen, der es ermöglicht, die Schritte intersubjektiv nachvollziehbar und transparent darzustellen. Die Schritte der Untersuchung können fortlaufend anhand der Anforderungen und Ziele des DBR-Ansatzes evaluiert und reflektiert werden. Aspekte, die es im Kontext des Forschungsvorhabens kritisch zu betrachten gilt, werden im Folgenden in Anlehnung an die Kritikpunkte im Kap. 3.1.6 kurz diskutiert.

Generalisierbarkeit: Die Ergebnisse können aufgrund eines kleinen Samples und des eher explorativen und weniger experimentellen Charakters nur bedingt generalisiert werden. Um den bereits oben genannten Kritikpunkte zu begegnen, wurde der DBR-Prozess durch klassische empirische Methoden ergänzt. Grundsätzlich wurde explorativ vorgegangen, wobei sich im dritten Zyklus – der

Hauptuntersuchung – des Mixed Method-Ansatzes bedient wurde. Es wurden qualitative Interviews, Usability-Studien und Befragungen durchgeführt. Die Interviews sollen Aufschluss über die Perspektive und Bedürfnisse der Nutzer und Nutzerinnen liefern.

Kausalattribution: Die Usability-Studien wurden durch randomisierte hypothesengeleitete Experimente ergänzt, um Aussagen hinsichtlich möglicher Kausalitäten treffen zu können. Dabei wurden mögliche Einflussfaktoren, soweit der Kontext dies zuließ, berücksichtigt.

Involviertheit von Forschenden und Praktikern/Praktikerinnen: Eine Schwierigkeit bestand allerdings in der Unkontrollierbarkeit bzw. nur bedingten Kontrollierbarkeit diverser Einflussfaktoren. Insbesondere bestand die Gefahr, dass die Personen während der Usability-Tests unterschiedlich stark von den Betreuern und Betreuerinnen unterstützt wurden. Um diesen Einfluss im Auswertungsprozess berücksichtigen zu können, wurden die Untersuchungen am Computer mit Video und die Interviews mit Audioaufnahmegeräten aufgezeichnet und entsprechend der wissenschaftlichen Gütekriterien analysiert. Zudem erfolgt im Anschluss an die Ergebnisdarstellung der Zyklen jeweils eine kritische Reflexion.

Narrativer Charakter: Das Risiko einer zu wenig ausführlichen Dokumentation konnte nur – wie auch in allen anderen wissenschaftlichen Studien – mit dem Anspruch an Wissenschaftlichkeit begegnet werden, indem der Prozess möglichst detailliert und intersubjektiv nachvollziehbar dargestellt wird. Des Weiteren wurden in jedem Evaluations-Zyklus die eingesetzten Methoden im Anschluss an die Ergebnisauswertung sowie in einer abschließenden Methodendiskussion reflektiert.

6 Zyklus 1: Paper Prototyping

Die erste Vorstudie wurde im Jahr 2010 durchgeführt. Zu dem Zeitpunkt lagen erste Konzept- und Layoutentwürfe für otu.lea vor. Vor der technischen Umsetzung der Inhalte für otu.lea wurde eine Entwicklungsumgebung – ein sogenanntes Framework (vgl. Kap. 7.7.1 und Abbildung 27) – realisiert. Die Entwicklungsumgebung ermöglicht eine einfache Erstellung von Items sowie deren Modifizierung. Zum Zeitpunkt der ersten Studie waren somit keine Aufgabeninhalte technisch umgesetzt. Allerdings galt es im Sinne des Usability-Prinzips „*Test early and often*“ (vgl. z. B. Krug, 2002; Nielsen, 2008), in einem möglichst frühen Entwicklungsstadium die ersten konzeptionellen Entwürfe zu testen. Daher wurde in der ersten Vorstudie ein Papierprototyp von otu.lea evaluiert. In Anlehnung an den DBR-Ansatz und deren Phasen (Gestaltung, Durchführung, Überprüfung, Re-Design) wird im Folgenden die erste Vorstudie aus dem Jahr 2010 dargestellt⁶⁶.

6.1 Zyklus 1: Gestaltung

Die Gestaltung des ersten Zyklus ist einerseits mit der Entwicklung und Definition des Evaluationsgegenstandes auf praktischer Ebene (1) und andererseits mit der Formulierung der Forschungsfrage (2) sowie der Methodenwahl und der Bestimmung des Untersuchungskontextes (3) auf theoretisch-methodischer Ebene anzusiedeln.

(1) Entwicklung und Definition des Evaluationsgegenstandes

Hinsichtlich der inhaltlichen Gestaltung wurden auf der Grundlage bisheriger Erkenntnisse über die Zielgruppe (vgl. Kap. 2.3), über die CLT (vgl. Kap. 2.4), über wahrnehmungspsychologische Aspekte (vgl. Kap. 2.6.3), und über Usability-Anforderungen (vgl. Kap. 2.6.4) sowie einer Sichtung und Systematisierung vorhandener computerbasierter Itemformate die Struktur für das Interface und erste Items entworfen. Diese berücksichtigten (soweit möglich) folgende Anforderungen.

Zielgruppe: Die Zielgruppe der funktionalen Analphabeten und Analphabettinnen zeichnet sich durch eine hohe Testangst (vgl. Döbert u. a., 2000), tendenziell geringe Computerkenntnisse (Niesyto 2009) und eine niedrige Literalität

⁶⁶ Ergebnisse des Paper Prototypings wurden bereits in Wolf und Koppel (2010) und Koppel (2011) veröffentlicht.

aus. So ist davon auszugehen, dass die Bedienkompetenz und Erfahrung mit Computern nicht sonderlich stark ausgeprägt sind.

Usability: Aufgrund fehlender Erkenntnisse über Usability-Anforderungen für funktionale Analphabeten und Analphabetinnen, generierten sich die Vorannahmen aus generellen Erkenntnissen zur Usability (vgl. Usability-Heuristiken Kap. 2.6.4) und Usability-Anforderungen für Personen mit eingeschränkten Fähigkeiten (Kap. 2.6.4.2). Da von einer tendenziell niedrigen ICT-Literacy auszugehen war, aber keine Studien über die Computerfähigkeiten von funktionalen Analphabeten und Analphabetinnen bekannt waren, wurde lediglich eine minimale Vertrautheit mit Maus und Tastatur vorausgesetzt. Es wurde auf eine einfache Interfacestruktur, auf einen iterativen Aufbau des Interfaces (um eine Überforderung der Teilnehmenden zu vermeiden) und auf die Konsistenz aktivierbarer Schaltflächen geachtet.

Cognitive Load Theory: Der CLT folgend wurden Annahmen darüber getroffen, wie der *Extraneous Load* möglichst gering beansprucht werden kann. Funktionen, wie das Blinken einer Schaltfläche, sollten die Nutzer-/Nutzerinnenführung sowie die Orientierung erleichtern. Zudem sollte eine Hilfefunktion implementiert werden, die bei Bedarf aufgerufen werden kann.

Für den ersten hier beschriebenen Zyklus wurden vier unterschiedliche Items aus den Dimensionen Lesen und Schreiben ausgewählt, die bereits von den entsprechenden Teilprojekten für die Papierversion entwickelt worden waren. Kriterien für die Auswahl der Items waren möglichst hohe Interaktivität, Variation der Itemformate sowie Items, bei denen Probleme in der Bedienbarkeit erwartet wurden. Somit wurden geschlossene Single Choice und halb offene Antwortformate ausgewählt, welche von den Probanden und Probandinnen sowohl das Anklicken einer Schaltfläche als auch die Eingabe von mehreren Buchstaben in einem Textfeld erforderten.

(2) Formulierung der Forschungsfragen

Mit dem ersten Zyklus sollten Antworten auf folgende Fragen ermittelt werden:

1. Können sich die Teilnehmenden orientieren?
2. Wissen sie, wie Maus und Tastatur zur Bearbeitung der Aufgaben zu nutzen sind?
3. Unterstützen simulierte Cues die Nutzer-/Nutzerinnenführung?

Die Kriterien für die Evaluation lieferten ausgewählte Heuristiken. Die Auswahlkriterien der Heuristiken waren Relevanz und Überprüfbarkeit. Es wurden die Heuristiken *Selbstbeschreibungsfähigkeit*, *Wahrnehmungssteuerung*, *Erwartungskonformität* sowie *Perspektivübernahme* identifiziert. Die elf übrigen Heuristiken wurden aus folgenden Gründen nicht als Evaluierungsgrundlage herangezogen: 1. Die konzeptionelle Entwicklung war noch nicht abgeschlossen,

wodurch die Heuristiken *Feedback*, *Individualisierbarkeit*, *Prozessangemessenheit*, *Joy of Use* sowie *interkulturelle Aspekte* nicht überprüft wurden (Auswahlkriterium Überprüfbarkeit). 2. Die technische Umsetzung war nicht realisiert, so dass keine Hinweise hinsichtlich der Heuristiken *Fehlertoleranz* und *Datensicherheit* gewonnen werden konnten (Auswahlkriterium Überprüfbarkeit). 3. Die Heuristiken *Konsistenz*, *Aufgabenangemessenheit* und *Steuerbarkeit* sollten erst in einem späteren Zyklus als Evaluationskriterium herangezogen werden.

Auf der theoretischen Grundlage wurden zur Überprüfung der ausgewählten jeweiligen Heuristiken Indikatoren formuliert:

Der Indikator für die *Selbstbeschreibungsfähigkeit* ist die Unterstützungsleistung, die von den Probanden und Probandinnen eingefordert wird und weder die Regeln des Paper Prototypings noch inhaltliche Aspekte der Literalität betreffen. Beispielsweise wenn die Probanden und Probandinnen Unterstützung benötigen, um das Experiment weiter durchführen zu können und sich die Unterstützungsleistung eindeutig auf Usability-Aspekte bezieht (z. B. wenn danach gefragt wird, wo der Buchstabe eingetragen werden soll oder wo die Aufgabe steht).

Die *Wahrnehmungssteuerung* wurde anhand des Einsatzes von Cues überprüft. Dabei galt es herauszufinden, ob Effekte dieser Unterstützungsfunktionen (in diesem Fall das simulierte Blinken einer Schaltfläche) auf die Reaktionszeit beobachtbar sind. Dafür wurden zwei Sequenzen mit unterschiedlichen experimentellen Bedingungen durchgeführt (s. Tabelle 12). Der Indikator für die Wahrnehmungssteuerung ist demnach die Zeit, die benötigt wird, um mit dem System zu interagieren. Es wurden dabei ausschließlich die Bereiche berücksichtigt, die nicht die Literalität betreffen, um einen Einfluss der Literalitätskompetenz (z. B. die Überlegungszeit zur Lösung einer Aufgabe) auszuschließen. Im Paper Prototyping-Experiment ist der Indikator lediglich auf die erste Sequenz anzuwenden, um einen möglichen Lerneffekt auszuschließen. Somit ist für die Wahrnehmungssteuerung anzeigend, wie viel Zeit die Probanden und Probandinnen benötigen, um das erste Lautsprechersymbol auf der ersten Seite anzuklicken.

Für die *Erwartungskonformität* ist anzeigend, ob die Probanden und Probandinnen grundsätzliche Funktionen zur Systeminteraktion beherrschen (z. B. das Anklicken einer Schaltfläche und das Eintragen von Buchstaben in ein Textfeld). Allerdings ist dieser Indikator nur bedingt überprüfbar, da ein papierbasiertes Instrument sich im Charakter von einem digitalen Instrument stark unterscheidet. So lässt sich in diesem Fall nur bedingt von Schwierigkeiten mit dem papierbasierten Instrument auf Schwierigkeiten mit einem digitalen System schließen.

Für die Verständlichkeit der Instruktionen ist ein Indikator die *Perspektivübernahme*. Wird durch die Interaktion ersichtlich, dass der Instruktionstext nicht verstanden wurde, ist das Maß an Verständlichkeit nicht ausreichend ist.

(3) Methodenwahl und Bestimmung des Untersuchungskontextes

Die Methodenwahl fiel auf ein *Paper Prototyping*-Experiment mit funktionalen Analphabeten und Analphabetinnen. Es sollte lediglich der Aufgabenbereich getestet werden. Weitere Bestandteile der Umgebung (wie z. B. der Login-Bereich) waren nicht Gegenstand des ersten Evaluationszyklus`.

Folgende Tabelle zeigt einen Überblick des Paper Prototyping-Experiments:

Zyklus 1	Vorstudie 1 Paper Prototyping			
N	5			
Heuristiken	Selbstbeschreibungsfähigkeit	Wahrnehmungssteuerung	Erwartungskonformität	Perspektivübernahme
Indikatoren	Eingeforderte Unterstützung	Benötigte Zeit zur Interaktion	Beherrschung grundsätzlicher Funktionen	Verständlichkeit der Instruktionen
Methoden	Paper Prototyping			

Tabelle 11: Übersicht Paper Prototyping-Experiment 2010

In den zwei durchgeführten Experimenten bestanden unterschiedliche experimentelle Bedingungen:

	Sequenz 1	Sequenz 2
Probanden-/Probandinnen -Anzahl	3	2
Cue	verbal	visuell
Hilfefunktion	nein	ja

Tabelle 12: Versuchsanordnung Paper Prototyping 2010

In der ersten Sequenz wurde keine Hilfefunktion angeboten. Das Blinken von Schaltflächen wurde nicht visuell simuliert, sondern es wurde lediglich verbal im Verlauf des Experiments darauf hingewiesen. In der zweiten Sequenz wurde der visuelle Hinweis in Form eines roten Kreises dargestellt. In Abhängigkeit der Aktion der/des Probanden/-in, wurde der rote Kreis platziert, um auf den Button hinzuweisen, der angeklickt werden muss und um der Reihenfolge entsprechend durch die Befragung zu gelangen. Die Hilfefunktion wurde in der zweiten Sequenz eingebunden. Diese war in Form eines Fotos dargestellt und wurde zu Beginn des Experiments den Teilnehmenden erläutert.

6.2 Zyklus 1: Durchführung

Der Kontakt zu den Probanden/-innen wurde über die ortsansässige Volkshochschule hergestellt, in der Alphabetisierungskurse stattfinden. Über die Kursleitenden wurden die Probanden und Probandinnen auf freiwilliger Basis akquiriert. Als Anreiz haben die Probanden und Probandinnen pro Person und Experimentteilnahme 15 Euro erhalten (Wolf & Koppel, 2010).

Das Paper Prototyping fand in zwei Sequenzen mit insgesamt fünf Probanden und Probandinnen statt. In der ersten Sequenz haben drei und in der zweiten Sequenz zwei Probanden und Probandinnen am Experiment teilgenommen⁶⁷. Dem Proband/Der Probandin wurden im Vorfeld die Personen, deren Rollen sowie die Spielregeln erläutert. Ihnen wurden der Umgang mit der Maus, der Tastatur und in der zweiten Sequenz die visuellen Cues erläutert. In der ersten Sequenz wurden die visuellen Cues nicht eingesetzt, sondern das Blinken einer Schaltfläche wurde nur verbal angemerkt. Zur Eingabe der Buchstaben wurde eine gängige Computertastatur eingesetzt. Folgendes Bild zeigt einen Probanden/eine Probandin aus der ersten Sequenz ohne Cues während er Bearbeitung einer Aufgabe mit Tastatureingabe.

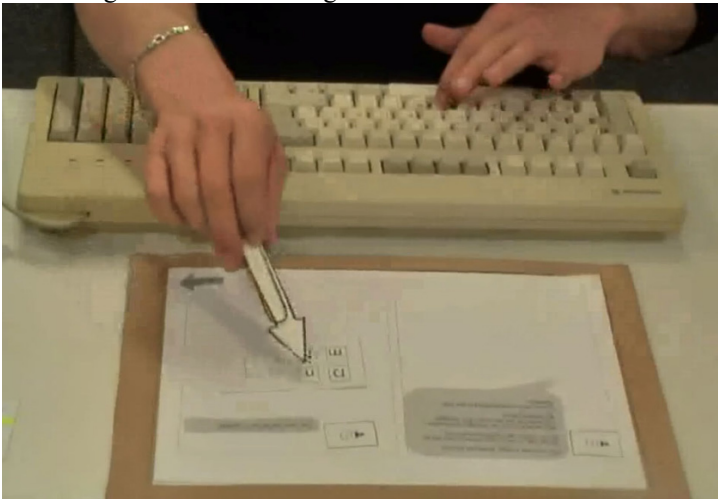


Abbildung 17: Paper Prototyping mit einem Papierprototyp von otu.lea, Sequenz 1

Ein Foto aus der zweiten Sequenz inklusive visuellen Cues mit einem Item aus der Dimension Schreiben zeigt das folgende Foto.

⁶⁷ Nielsen (2006) empfiehlt, ein Paper Prototyping-Experiment mit 4-7 Probanden/-innen durchzuführen.

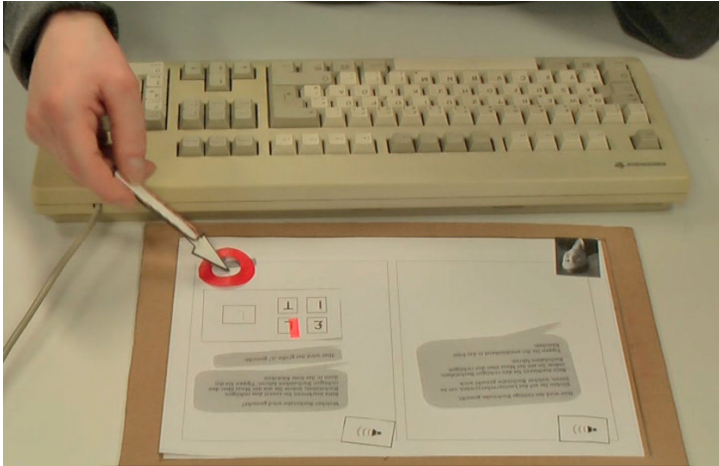


Abbildung 18: Paper Prototyping mit einem Papierprototyp von otu.lea, Sequenz 2

Das Experiment wurde auf Video aufgezeichnet. Es wurden nur die Hände der Teilnehmenden gefilmt, um deren Anonymität zu wahren. Im Anschluss wurden den Probanden/-innen Fragen zu deren Empfindungen, Schwierigkeiten und Verbesserungsvorschlägen gestellt.

6.3 Zyklus 1: Analyse

Für die Analyse wurden sowohl die Interviews als auch die Videos herangezogen.

1. Datenauswertung

Das Datenmaterial wurde nach der qualitativen Inhaltsanalyse ausgewertet (eine detaillierte Beschreibung der qualitativen Inhaltsanalyse erfolgt in Kap. 8.2.4). Der Fokus lag auf dem Antwortverhalten der Probanden und Probandinnen.

2. Ergebnisdarstellung und -interpretation

In folgender Tabelle sind die Ergebnisse des Paper Prototyping-Experiments zusammengefasst.⁶⁸ Die Probandinnen/Probanden werden anonymisiert mit Buchstaben und der Ziffer „1“ bezeichnend für Zyklus „1“ betitelt.

⁶⁸ Die Ergebnisse wurden in ähnlicher Form in Wolf & Koppel (2010) und Koppel veröffentlicht.

Heuristik und Indikator	Beobachtung
Selbstbeschreibungsfähigkeit: Externe Unterstützung	Eingeforderte Unterstützung: Von Proband/Probandin C1 wurde Unterstützung eingefordert, indem er/sie äußerte nicht zu wissen, was er/sie machen solle. Er/Sie fährt allerdings ohne weitere Unterstützung fort. <i>Abgesehen von Proband/Probandin C1 wurde kein weiterer Unterstützungsbedarf formuliert.</i> Nutzung der Hilfefunktion (die Hilfefunktion war in der ersten Sequenz (Proband/Probandin A1-C1) nicht implementiert); Proband/Probandin E1 nutzt die Hilfefunktion dreimal. 1) Das erste Mal als scheinbar unklar ist, wie er/sie mit der Aufgabenbearbeitung fortfahren kann. Zum Zeitpunkt der Funktionsnutzung sind nur Lautsprechersymbole zu sehen. Nachdem die Person die Hilfefunktion aktiviert und ihr zugehört hat, fährt sie unmittelbar ohne weitere Verzögerung mit der Aufgabenbearbeitung fort. 2) Bei der zweiten Nutzung der Hilfefunktion benötigt der Proband/die Probandin Unterstützung, weil er/sie anscheinend nicht weiß, wo die richtige Antwort anzuklicken ist. Die Hilfefunktion liefert nur generelle Informationen und ist nicht auf die Aufgabe bezogen, so dass die Person nur durch willkürliches Versuchen die richtige Schaltfläche aktiviert. 3) Ein weiteres Mal aktiviert Proband/Probandin E1 die Hilfefunktion, als bei der Aufgabe „Simones Bestellung“ der Soundbutton zu aktivieren ist um zu erfahren, welche Buchstaben einzutragen sind. Allerdings hat die Funktionsnutzung keinen positiven Effekt: die Person aktiviert auch anschließend nicht die Lautsprecher, um sich die Aufgabe ein weiteres Mal vorlesen zu lassen. Durch Probieren gelingt es ihr allerdings, die Funktionen zur Lösung der Aufgabe aufzurufen. Proband/Probandin D1 hat die Hilfefunktion nicht genutzt. Proband/Probandin E nutzen die Hilfefunktion nicht.
Wahrnehmungssteuerung: Zeit	Effekte der visuellen Cues: Hinsichtlich der visuellen Hinweise (Cues) konnte festgestellt werden, dass die Probanden und Probandinnen A1-C1 in der ersten Sequenz – in der verbal auf das Blinken einer Schaltfläche hingewiesen wurde – mehr Bedenkzeit als in der zweiten Sequenz (Proband/Probandin D1 und E1) benötigen (5 Sekunden bis zu über einer Minute), um den nächsten Schritt – das Anklicken des Soundbuttons – durchzuführen. In der zweiten Sequenz benötigen die Probanden und Probandinnen 1-2 Sekunden zum

Heuristik und Indikator	Beobachtung
	Anklicken des Lautsprechersymbols.
<i>Erwartungskonformität:</i> Beherrschung grundsätzlicher Funktionen	Dem Proband/Der Probandin B1 scheint die Funktion der Hochsteltaste nicht bekannt, da er/sie bei der Aufgabe „Buchstabensuche“ die Buchstaben klein schreibt, obwohl in der Aufgabenstellung explizit auf Groß- bzw. Kleinschreibung hingewiesen wird. Bei Proband/Probandin C1 ist die Beherrschung grundsätzlicher Funktionen nicht erschöpfend zu beantworten, da diese Person lediglich die Maus nutzt. Die Probanden und Probandinnen A1, D1 und E1 scheinen die grundsätzlichen Funktionen zu beherrschen. Keiner der Testteilnehmenden klickt mit der Maus in die Lücke, bevor er/sie etwas einträgt.
<i>Perspektivübernahme:</i> Verständlichkeit der Instruktionen	Aufgabe Kinoflyer: Proband/Probandin C1 scheint die Instruktionen zum Teil nicht zu verstehen: Die Antwortmöglichkeiten werden nicht mit dem Stimulus (in diesem Fall dem Kinoflyer) in Verbindung gebracht, sondern die Person versucht, die Frage mit Allgemeinwissen zu beantworten: die Person glaubt, dass dienstags die Kinos geschlossen haben. Aufgabe Buchstabensuche: Bei der Aufgabe „Buchstabensuche“ tippt Proband/Probandin C1 nicht die Buchstaben in die Kästchen, bei der Aufgabe „Simones Bestellung“ buchstabiert er/sie zwar die Wörter, trägt sie aber nicht mit der Tastatur ein. Dem Proband/Der Probandin D1 waren die Spielregeln teilweise unklar: beispielsweise dass die Buchstaben mit der Tastatur eingetippt werden müssen. Proband/Probandin E1 vergisst bei der Aufgabe „Buchstabensuche“ den Buchstaben in das freie Kästchen einzutragen. Des Weiteren vergisst er/sie teilweise, die Zahlen bei der Aufgabe „Simones Bestellung“ einzutragen. Die Instruktionen wurden somit nicht immer in vollem Umfang berücksichtigt.

Tabelle 13: Ergebnisse des Paper Prototyping 2010

Des Weiteren wurde beobachtet, dass die Probanden und Probandinnen während des Experiments nervös waren. Dies äußerte sich in zitternden Händen und Schweißausbrüchen.

Für die Interpretation der Ergebnisse werden die Beobachtungen mit den in diesem Experiment fokussierten Heuristiken in Beziehung gesetzt.

Selbstbeschreibungsfähigkeit: Nur ein Proband/eine Probandin formulierte Unterstützungsbedarf, fuhr jedoch ohne Unterstützung fort. Die Hilfefunktion war nur in der zweiten Sequenz und somit nur für die Teilnehmenden D1 und E1 nutzbar. Eine Person aktivierte zu drei unterschiedlichen Zeitpunkten die Hilfefunktion. Dass ein Proband/eine Probandin die Hilfefunktion trotz zwischenzeitlichem Bedarf (siehe Ergebnisse Perspektivübernahme) nicht nutzte, ist wahrscheinlich entweder darin begründet, dass die Person nach kurzer Zeit selbst herausfand, wie die Aufgabe zu lösen ist oder, dass sie vergessen hat, dass eine Hilfefunktion genutzt werden kann.

Wahrnehmungssteuerung: Auf der Grundlage der CLT ist zu vermuten, dass der *Extraneous Load* stärker ausgelastet ist, wenn die Cues lediglich verbal erläutert werden. Die Beobachtungen bestätigen diese Vermutung: visuelle Cues scheinen stärker als verbale Cues die Orientierung zu erleichtern. Begründet ist dies womöglich darin, dass die verbale Erläuterung nur kurzzeitig – während der Erläuterung – präsent ist und nach der Erläuterung wieder vergessen wird. Auch kann ein rein verbaler Hinweis einen geringeren Aufforderungscharakter haben, als ein visueller Hinweis in Form eines großen roten Kreises, der sich farblich stark von dem Hintergrund abhebt⁶⁹. Auch wenn diese Beobachtungen keine empirische Schlussfolgerung zulassen kann vermutet werden, dass sich die Bearbeitungszeit und die Auslastung des *Extraneous Load* mit visuellen Cues reduzieren lassen.

Erwartungskonformität: Das Experiment liefert Hinweise darauf, dass der Umgang mit Tastatur und Maus für die Teilnehmenden keine Selbstverständlichkeit ist. Einerseits wurde dies durch die Bearbeitungszeit deutlich, andererseits war den Teilnehmenden nicht immer ersichtlich, wann die Maus zu nutzen ist und wo der Buchstabe auf dem Monitor erscheint, der auf der Tastatur eingegeben wurde. In allen Fällen scheinen die Teilnehmenden nicht zu wissen, dass sie mit der Maus erst in eine Lücke klicken müssen, bevor sie etwas eintragen können. Dies ist vermutlich in den Bedingungen des Experiments begründet: Paper Prototyping erfordert eine Abstraktions- und Transferleistung, nämlich sich vorzustellen und so zu handeln, als ob man vor einem Computer sitzen würde, obwohl dies nicht der Fall ist. Dadurch, dass drei Personen die Hochsteltaste korrekt nutzen ist davon auszugehen, dass sie Erfahrungen mit dem Computer haben und womöglich auch wissen, dass sie mit der Maus den Cursor in dem Textfeld platzieren müssen, um etwas eintippen zu können. Es kann daher nicht

⁶⁹ Die verbale Erläuterung stellte allerdings keine reale Option für die Implementation in *otu.lea* dar. Die experimentelle Bedingung wurde als verbale Äußerung eingefügt, um eine Vergleichsmöglichkeit zum visuellen Hinweissigna zu haben. Es wurde sich für eine Erläuterung entschieden, da vermutet wurde, dass sich die Probanden und Probandinnen ohne verbale Hinweise nur schwer hätten orientieren können und die Frustration gestiegen wäre.

beantwortet werden, ob die Personen die grundsätzlichen Funktionen hinsichtlich der Bedienung von otu.lea beherrschen.

Perspektivübernahme: Für drei Personen (C1, D1, E1) lässt sich aufgrund der nicht berücksichtigten Instruktionen (z. B. dass bei der Aufgabe „Buchstabensuche“ neben dem Anklicken des richtigen Buchstabens auch der Buchstabe eingetippt werden muss) vermuten, dass die Instruktionen nicht verständlich genug sind. Möglicherweise sind die Instruktionstexte zu lang. Bei (halb)offenen Aufgabenformaten müssen sich die Nutzer und Nutzerinnen merken, dass sie erstens etwas anklicken müssen, um sich das einzutragende Wort vorlesen zu lassen. Zweitens, dass sie mit der Maus in die Lücke klicken und drittens mit der Tastatur die Buchstaben eintragen müssen. Aber auch hier kann die Ursache (wie auch bei Erwartungskonformität erläutert) darin liegen, dass ein Paper Prototyping-Experiment eine Abstraktions- und Transferleistung erfordert. Es kann damit auch hier keine klare Schlussfolgerung gezogen werden.

Eine methodische Herausforderung im Paper Prototyping bestand in der geforderten Abstraktions- und Transferleistung seitens der Probanden und Probandinnen. Diese mussten während des Experiments abstrahieren, dass sie die Lösungen zwar mit der Tastatur eingeben oder mit der Maus angeklickt wurden, diese aber nicht auf einem Monitor, sondern auf einem ihn vorliegenden Papier erschienen, welches lediglich den Monitor simulierte. Dies führte anfangs zu Irritationen seitens der Probanden und Probandinnen (zum Beispiel der Versuch, die Lösung direkt auf dem „Monitor“ einzutragen). In darauf folgenden Aufgaben trat die Irritation allerdings nicht mehr auf. Es fand vermutlich eine Gewöhnung und eine Art Lerneffekt bzw. Verinnerlichung der Spielregeln statt. Zusammenfassend lieferte das Experiment über das Nutzer-/Nutzerinnenverhalten Hinweise für die weitere Gestaltung. Schließlich trug das Paper Prototyping-Experiment dazu bei, Erfahrungen über die Zielgruppe zu sammeln und diese in weiteren Entwicklungsschritten zu berücksichtigen, wie es sowohl die Usability-Evaluation als auch der DBR-Ansatz vorsehen.

6.4 Zyklus 1: Re-Design

Das Re-Design wurde sowohl auf der Grundlage der Beobachtungen aus dem Paper Prototyping als auch unter Berücksichtigung weiterer Entwicklungsideen durchgeführt.

Heuristik	Modifizierungsvorschläge für otu.lea.
Selbstbeschreibungsfähigkeit	Die Hilfefunktion ist zu Beginn eingehend zu erläutern. Zudem ist die Hilfefunktion kontextsensitiv (abgestimmt auf die aktuelle Umgebung/Seite) zu gestalten, so dass die Nutzer und Nutzerinnen Unterstützung für die Funktionen und Aufgaben erhalten, die sie momentan bearbeiten.
Wahrnehmungssteuerung	Visuelle Cues sind einzusetzen, um die Orientierung zu erleichtern.
Erwartungskonformität	Bei (halb)offenen Aufgabenformaten ist den Teilnehmenden zu erläutern, dass sie erst in eine Lücke klicken müssen, um anschließend etwas eintragen zu können.
Perspektivübernahme	Die Instruktionstexte werden mit kürzeren Sätzen versehen und es wird darauf hingewiesen, dass der Text mehrfach abgespielt werden kann. Zudem werden Möglichkeiten geprüft, die Instruktionen zu reduzieren oder aufzuteilen.

Tabelle 14: Modifikationen von otu.lea auf der Grundlage des ersten Zyklus

In der weiteren Design-Entwicklung wurden auf Grundlage der Ergebnisse insbesondere die Animation der Schaltflächen und die Ausweitung der Hilfestellung berücksichtigt.

Die Schaltflächen wurden in logischer Abfolge animiert. Da die Hilfefunktion wenig genutzt wurde, wurde ein Tutorial implementiert, welches zu Beginn auf die Hilfefunktion hinweist. Daraus entwickelte sich zudem die Idee, differenzierte Hilfefunktionen einzubinden, die sich auf verschiedene Bereiche und somit auch Funktionen beziehen. Ist eine Aufgabe im Single oder Multiple Choice-Format zu bearbeiten, erläutert die Hilfefunktion auf der entsprechenden Seite die Funktionsweise dieses Itemformats. Wird die Bearbeitung einer Freitext-Aufgabe gefordert, können über die Hilfefunktion die entsprechenden Informationen zur Funktionsweise des Multiple Choice-Formats abgerufen werden.

6.5 Reflexion des Paper Prototyping-Experiments

Die Effekte der visuellen Cues wären vermutlich stärker ausgefallen, wenn in einer Sequenz keine Hinweise (weder verbal noch nonverbal) und in der anderen Sequenz der nonverbale visuelle Hinweis eingesetzt worden wären. Es ist jedoch festzuhalten, dass das Paper Prototyping aufgrund der niedrigen Probanden/Probandinnen-Anzahl keine Ergebnisse liefert, die empirische Rückschlüsse zulassen. Um aussagekräftigere Ergebnisse zu erzielen, wurden diese Überle-

gungen bezüglich der experimentellen Bedingung in der Hauptuntersuchung (s. Kap. 8) berücksichtigt.

Nicht alle Probanden und Probandinnen interagierten zu Beginn des Experiments spielregelkonform. Es wurden zwar teilweise die richtigen Lösungen formuliert, doch tippten sie diese nicht auf der Tastatur ein. Dies könnte darin begründet sein, dass den Probanden/Probandinnen der Transfer zwischen papierbasiertem Interface und der gedanklichen Übertragung des Effekts der eigenen Handlung auf einen Bildschirm schwer fiel. Im Verlauf des Experiments wurden die Spielregeln aber zunehmend berücksichtigt. Zudem könnten die Videokamera und die weiteren anwesenden Teilnehmenden des Experiments zur Steigerung der Nervosität beigetragen haben. Um den Probanden und Probandinnen die Möglichkeit zu geben, die Spielregeln zu verinnerlichen und Nervosität abzubauen kann eine Übungssequenz im Vorfeld hilfreich sein. Einerseits trägt dies zur Verständlichkeit und Umsetzung der Bedingungen bei. Andererseits stärkt eine Übungssequenz das Vertrauen zwischen Probanden/Probandinnen und den weiteren Personen.

7 Zyklus 2: Usability-Studie 2012

Die zweite Vorstudie fand im Jahr 2012 an einer Volkshochschule in Norddeutschland statt. Ein Großteil der Aufgaben der Online-Testumgebung war programmiert und funktionsfähig. Die automatisierte Auswertung und damit auch die Rückmeldungen waren noch nicht eingebunden. Aussagekräftige und umfangreiche Ergebnisse versprachen die Durchführung von Usability-Tests mit der Testumgebung und realen Nutzern und Nutzerinnen sowie die Durchführung von Interviews.

7.1 Zyklus 2: Gestaltung

Wie auch im ersten Zyklus gliedert sich die Gestaltungsphase in (1) Entwicklung und Definition des Evaluationsgegenstandes, (2) Formulierung der Forschungsfrage und (3) Methodenwahl und Bestimmung des Untersuchungskontextes gegliedert.

Während der ersten Programmierphase (2010 bis 2011) fanden regelmäßige informelle produktzentrierte Heuristische Evaluationen mit Experten und Expertinnen statt. Ein erster Funktionsfähiger Prototyp wurde nun mit potenziellen Nutzern und Nutzerinnen getestet.

(1) Entwicklung und Definition des Evaluationsgegenstands

Auf der inhaltlichen und praktischen Ebene wurden in der weiteren Entwicklung der Online-Testumgebung auf der Basis der theoretischen Vorannahmen zur CLT und zur Usability die Ergebnisse des Paper Prototypings berücksichtigt. Es wurde ein erster Prototyp technisch – wie bereits in der Phase des Re-Designs im ersten Zyklus beschrieben – umgesetzt, in welchem Cues sowie eine rudimentäre Hilfefunktion implementiert worden sind. Die zweigeteilte Struktur des Interfaces wurde beibehalten. Zudem wurden neben den inhaltlich an das Storyboard angelehnten Instruktionstexten auch weitere Elemente des Storyboards in Form von Bildern und zusätzlichen Informationen zu den Charakteren eingebunden. Kriterium für die Auswahl von Aufgaben für den zweiten Zyklus war die Funktionalität. Da ein Teil der Aufgaben nicht funktionsfähig waren, konnten nicht alle Aufgaben für die zweite Studie zur Evaluation zur Verfügung gestellt werden. Den Probanden und Probandinnen standen dennoch aus allen Dimensionen Aufgaben zur Verfügung. Die Aufgaben wurden den Niveaus „einfach“, „mittel“ und „schwierig“ zugeordnet. Entsprechend der Selbsteinschätzung der Probanden und Probandinnen konnten sie das Niveau wählen.

(2) Formulierung der Forschungsfragen

Ziel der zweiten Vorstudie war herauszufinden, ob die Teilnehmenden das Programm dem Zweck entsprechend nutzen können und wie sie das Programm bewerten. Um dieses Ziel zu erreichen, sollten folgende Fragen beantwortet werden:

1. Welche und wie viel Unterstützung benötigen die Teilnehmenden bei der Bearbeitung der Aufgaben?
2. Wie zufrieden sind die Teilnehmenden mit dem Programm?

Die Evaluationskriterien wurden auf Basis ausgewählter Heuristiken festgelegt und konkretisiert. Für die Auswahl der Heuristiken wurden – wie auch im Paper Prototyping – die Kriterien Überprüfung und Relevanz angelegt.

Es wurden die Heuristiken *Erwartungskonformität*, *Selbstbeschreibungsfähigkeit*, *Perspektivübernahme*, *Erlernbarkeit* und *Joy of Use* identifiziert. Im Vergleich zum Paper Prototyping-Experiment wurde die Heuristik *Wahrnehmungssteuerung* in dieser Untersuchung nicht berücksichtigt, da die Relevanz dieser sich aus den experimentellen Bedingungen – der Einsatz bzw. das nicht Einsetzen von Cues – des Paper Prototypings ergab. Es wurden zwar auch in diesem Usability-Tests Cues eingesetzt, jedoch nicht systematisch, so dass ein Effekt nicht untersucht worden ist.

Indikatoren für die Heuristik *Erwartungskonformität* ist die Beherrschung grundsätzlicher Funktionen, um mit dem Computer interagieren zu können: Umgang mit der Maus sowie die Fähigkeiten, Buchstaben in ein Textfeld einzutippen und die Hochstelltaste für die Großschreibung zu nutzen.

Der Indikator für die Heuristiken *Selbstbeschreibungsfähigkeit* ist, wie auch im vorherigen Zyklus, die eingeforderte Unterstützung durch die Hilfefunktion und/oder die betreuenden Personen. Hierbei wurden die Probanden und Probandinnen drei Ausprägungen zugeordnet: Proband/Probandin hat „viel Unterstützung“ eingefordert, Proband/Probandin hat „wenig Unterstützung“ eingefordert und Proband/Probandin hat „keine Unterstützung“ eingefordert. Hierbei bedeutet „viel Unterstützung“, dass der Proband/die Probandin an mindestens zehn unterschiedlichen Stellen Unterstützung eingefordert hat; wenn die Person an weniger als zehn unterschiedlichen Stellen Unterstützung eingefordert hat, wird sie der Ausprägung b) „wenig Unterstützung“ zugeordnet. Zu kontrollieren war zum einen, in wie weit der Proband/die Probandin tatsächlich Unterstützung benötigte. Möglich ist, dass eine Person die Funktionen nicht versteht und dennoch etwas anklickt oder in Textfelder einträgt. Eine Kontrollmöglichkeit ist, die Zeit zur Beantwortung der Fragen zu berücksichtigen. Hat eine Person nur sehr wenig Zeit benötigt, um die Fragen zu beantworten, kann davon ausgegangen werden, dass sie die Aufgaben nicht inhaltlich bearbeitet hat. Zum anderen ist möglich, dass ein Proband/eine Probandin den Betreuer/die Betreuerin um Unterstützung gebeten hat, obwohl der Proband/die Probandin es auch ohne Unterstützung geschafft hätte und beispielsweise dachte: „Wenn die Personen schon hier sind,

kann ich sie auch fragen, dann muss ich nicht selbst nach einer Lösung suchen." Um dies zu verhindern wurden die Betreuer und Betreuerinnen instruiert, die Probanden und Probandinnen in einer eigenständigen Lösungsfindung zu unterstützen.

Für die *Perspektivübernahme* ist die Verständlichkeit der Instruktionen hinweisend.

Aussagen von den Teilnehmenden zur Orientierung, zur Informationsangemessenheit sowie zu Schwierigkeiten und Problemen sind Indikatoren für *Erlernbarkeit*.

Für die Heuristik *Joy of Use* werden Aussagen zur Zufriedenheit der Nutzer und Nutzerinnen und deren Veränderungsideen herangezogen.

(3) Methodenwahl und Bestimmung des Untersuchungskontextes

Für die Überprüfung des Prototyps wurden mit dem Usability-Test ein interaktionszentrierter und mit den Interviews ein benutzer-/benutzerinnenzentrierter Messansatz verwendet.

Zyklus 2	Vorstudie 2: Usability-Studie				
N	11				
Heuristiken	Erwartungskonformität	Selbstbeschreibungsfähigkeit	Perspektivübernahme	Erlernbarkeit	Joy of Use
Indikatoren	Beherrschung grundsätzlicher Funktionen	Eingeforderte Unterstützung	Verständlichkeit der Programminstruktionen	Aussagen zu Orientierung, Informationsangemessenheit, Schwierigkeiten/Problemen	Zufriedenheit, Veränderungsideen
Datenmaterial	Videoanalyse	Videoanalyse	Videoanalyse, Interview	Interview	Interview

Tabelle 15: Übersicht Usability-Studie 2012

Es nahmen elf Probanden und Probandinnen aus zwei Alphabetisierungskursen an einer Volkshochschule an der Usability-Studie teil. Die Teilnahme war freiwillig. Die Personen waren zwischen 19 und 64 Jahren alt, sechs davon waren weiblich und fünf männlich. Es wurden mit jedem Proband/jeder Probandin ein Usability-Test mit Videoaufzeichnung und ein fokussiertes Interview durchgeführt.

7.2 Zyklus 2: Durchführung

Die Studien wurden im Rahmen von zwei Alphabetisierungskursen durchgeführt. Vor dem Usability-Test wurden die Ziele der Studie sowie der Projektkontext erläutert und die mitwirkenden Personen vorgestellt.

Jeder teilnehmenden Person wurde ein Betreuer/eine Betreuerin zugewiesen. Die betreuende Person war für die Videoaufzeichnung und bei Bedarf zur Unterstützung für die Probanden und Probandinnen zuständig. Die Betreuungspersonen wurden im Vorfeld instruiert, sich möglichst im Hintergrund aufzuhalten und nur einzugreifen und/oder zu unterstützen, wenn der Verlauf des Tests gefährdet war. Bezüglich der Zielgruppe war von einer großen Testangst auszugehen und es lag im Ermessen der betreuenden Person, wann einzugreifen war, um die Probanden und Probandinnen nicht zu überfordern bzw. um deren Teilnahmemotivation aufrecht zu halten.

Bei dem Usability-Test standen alle Dimensionen (Lesen, Schreiben, Mathematik und Sprachempfindung) zur Auswahl. Die Betreuer und Betreuerinnen wurden zu Beginn instruiert, mit welcher Dimension zu beginnen ist, so dass für den ersten Durchlauf die Dimension vorgegeben war. Dies diente der Gewährleistung, dass alle Dimensionen von mindestens zwei Personen getestet wurden. Über eine Selbsteinschätzung konnten die Teilnehmenden „einfach“, „mittel“ und „schwer“ auswählen; in der Dimension Sprache konnte lediglich das Niveau „einfach“ angewählt werden. Wurde eine Dimension in dem selbst gewählten Schwierigkeitslevel absolviert, konnten die Teilnehmenden die weitere Dimension frei wählen. Ziel war, dass die Teilnehmenden zwei Dimensionen bearbeiten. Unabhängig des Testfortschritts sollte der Test nach ca. 20 Minuten beendet werden. Anschließend wurden die Interviews in einem separaten Raum durchgeführt.

7.3 Zyklus 2: Analyse

Der Fokus der Datenanalyse lag auf dem Interviewmaterial. Die Videos wurden für die Beobachtung des Antwortverhaltens, des Umgangs mit dem Computer sowie der Nutzung der Hilfefunktion herangezogen.

1. Datenauswertung

Das Videomaterial und die Interviews wurden mit der qualitativen Inhaltsanalyse nach Mayring und Brunner (2010) ausgewertet (vgl. auch Kap. 8.2.4). Auf der Grundlage theoretischer Erkenntnisse und insbesondere der Usability-Heuristiken wurde ein Auswertungsleitfaden entwickelt. Die Interviews wurden transkribiert und entsprechend der qualitativen Inhaltsanalyse codiert.

2. Ergebnisdarstellung und -interpretation

Die Usability-Tests dauerten ca. 20 Minuten. Zwei Tests stellten eine Ausnahme dar und dauerten jeweils 46 Minuten. Die Interviews dauerten ca. 15 Minuten. Den Teilnehmenden wurde von der Online-Testumgebung ein fünfstelliger Code aus Großbuchstaben und Zahlen zugewiesen. Einen exemplarischen Ausschnitt des Usability-Tests stellt folgendes Bild dar:

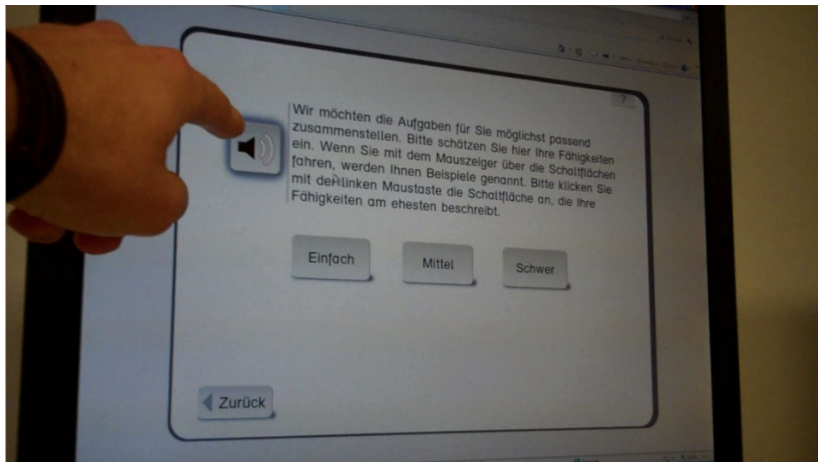


Abbildung 19: Exemplarischer Ausschnitt aus dem Usability-Test 2012

Das Foto zeigt den Monitor während der Durchführung des Usability-Tests. Der Proband/Die Probandin versucht, selbstständig den Text zu lesen. Daraufhin weist der Betreuer den Probanden/die Probandin auf die Funktion des Lautsprechers hin.

In der folgenden Tabelle sind die Ergebnisse zusammengefasst⁷⁰:

	Heuristik und Indikator(en)	Beobachtung
Videoauswertung (n=11)	Selbstbeschreibungsfähigkeit: Eingeforderte Unterstützung	Eingeforderte Unterstützung: Fünf (A2, B2, C2, D2) von zehn Probanden und Probandinnen haben viel, sechs Personen (E2, F2, G2, H2, I2, J2) wenig Unterstützung eingefordert. Zwei Personen haben lediglich am Anfang bei der Codeeingabe und der Niveauswahl Unterstützung benötigt. Bei einer Person (J2) hat der Betreuer/die Betreuerin sehr schnell eingegriffen, so dass nicht beurteilbar ist, wie viel Unterstützung sie benötigt hätte. Das Video ist hinsichtlich der Heuristik Selbstbeschreibungsfähigkeit nicht bewertbar. Keine Person hat den Usability-Test ohne Unterstützung absolviert. Login-Sequenz: Acht Probanden und Probandinnen klickten auf die Schaltfläche „Ich habe bereits einen Code“. Zwei haben ohne Unterstützung die Möglichkeit genutzt, sich einen Code erneut zuweisen zu lassen. Sieben Personen benötigten Unterstützung von den Betreuern und Betreuerinnen bei der Anmeldung. Eine Person (C2) wurde von dem Betreuer/der Betreuerin direkt darauf hingewiesen, dass sie noch keinen Code besitzt. Eine Person (J2) hat die richtige Schaltfläche ohne Unterstützung angeklickt.
	Erwartungskonformität: Beherrschung grundsätzlicher Funktionen	Alle Probanden und Probandinnen konnten ohne Unterstützung Buchstaben und Wörter in Textfelder eintragen. Auch das Schreiben von Großbuchstaben wurde beherrscht. Das Anklicken von Schaltflächen mit der Maus stellte für keine teilnehmende Person ein Problem dar. Die grundsätzlichen Funktionen beherrschten somit alle Probanden und Probandinnen.
	Perspektivübernahme (Interviewauswertung): Verständlichkeit der Instruktionen	Aufgabe Buchstabensuche: Bei der Schreibaufgabe „Buchstabensuche“ vergaß ein Proband (H2) ab dem 4. Item (von 8) das Anklicken des richtigen Buchstabens und hat nur den Buchstaben in das freie Feld eingetragen. Die Person J2 vergaß beim ersten Item das Anklicken, beim zweiten Item das Eintragen des Buchstabens. Die Person wurde allerdings unmittelbar darauf aufmerksam gemacht, so dass sie letztlich bei allen Items einen Buchstaben anklickte und einen Buchstaben eintrug. Simones Bestellung: Bei der Schreibaufgabe „Simones Bestellung“ mussten in separate Felder Ziffern und Silben eingetragen werden. Eine Person (K2) hat keine Zahlen eingetragen.

⁷⁰ Die Probandinnen/Probanden werden mit den Buchstaben A-M und der Ziffer „2“ (für Zyklus „2“) anonymisiert benannt.

	Heuristik und Indikator(en)	Beobachtung
Interviewauswertung (n=11)	Perspektivübernahme (Videoanalyse): Verständlichkeit der Instruktionen	Zehn Probanden und Probandinnen haben gesagt, sie hätten die Instruktionen inhaltlich verstanden. Ein Person sagte, der Umfang und die Anzahl der Texte sei „ein bisschen viel“.
	Erlernbarkeit: Aussagen zu Orientierung, Informationsangemessenheit, Schwierigkeiten, Problemen	Für sieben Teilnehmende erschien die Nutzung anfangs schwierig, am Ende jedoch leicht. Zwei Personen sagten aus, sich zu orientieren sei am Anfang genau so schwierig wie am Ende gewesen. Abgesehen von einer Person haben die Teilnehmenden ausgesagt, sich gut orientieren zu können. Eine Person erwähnte, dass es für sie schwierig gewesen sei, denn es müsse erst mal erkannt werden, was die dargestellten Symbole und Grafiken zu bedeuten haben. Schwierigkeiten wurden von zwei Personen bezüglich der Codeeingabe genannt.
	Joy of Use: Zufriedenheit, Veränderungsideen	Sechs Personen hat das Programm sehr gut gefallen, vier Personen fanden das Programm gut, eine Person mittelmäßig und eine Person war nicht zufrieden. Sie gab als Grund an, die Aufgabenlösungen und der Kick haben ihr gefehlt. Hinsichtlich des Storyboard äußerten sich sieben Personen positiv, zwei Personen indifferent und eine Person negativ. Als negativer Aspekt wurde die Größe der Bilder genannt: „Wir sind erwachsene Leute, da muss man nicht irgendwie so einen Kinderkram machen.“ (A2) Als Veränderungsideen wurden der Wunsch nach mehr Videos, mehr Geschichten und kürzeren Texte genannt.
k		

Tabelle 16: Ergebnisse der Usability-Studie 2012

Die Ergebnisse werden im folgenden Abschnitt bezogen auf die Heuristiken interpretiert.

Selbstbeschreibungsfähigkeit: Die Probanden und Probandinnen forderten insbesondere dann Unterstützung ein, als sie sich einloggen sollten. Auf der ersten Seite wurden die Probanden und Probandinnen aufgefordert anzugeben, ob sie bereits teilgenommen haben. Wenn sie auf „ja“ klickten, wurden sie auf der darauf folgenden Seite gebeten, ihren Code einzugeben. Haben die Probanden und Probandinnen „nein“ gewählt, wurde ihnen auf der folgenden Seite ein Code angezeigt, den sie in die freien Felder eintragen sollten. Die Informationen

wurden jeweils auditiv und visuell bereitgestellt. Alle Teilnehmer und Teilnehmerinnen wählten „ja“, obwohl sie zum ersten Mal teilnahmen. Es werden zwei Gründe für das Verhalten vermutet:

Vermutung 1: Die Probanden und Probandinnen verwechseln otu.lea mit dem bereits bestehenden Lernprogramm ich-will-lernen.de. Hinweise darauf liefern die Videos, in denen kommentiert wird, dass sie einen Code für ein anderes Programm bekommen hätten.

Vermutung 2: Der Instruktionstext ist zu lang und nicht verständlich.

Um die Vermutungen zu klären, hätten eine kommunikativen Validierung (vgl. z. B. Seale, 1999) und/oder weitere Studien mit der *Thinking Aloud*-Methode Aufschluss gegeben, in der auch während der Testdurchführung Rückfragen über die Gründe der Aktionen gestellt werden. Die Thinking Aloud-Methode wurde für die Zielgruppe von den Evaluationsexperten/-expertinnen jedoch als inadäquat bewertet, da die parallele Ausführung von Aufgaben und verbaler Beschreibung bzw. Reflexion durch die Probanden und Probandinnen zu einer Überforderung geführt hätten. Eine kommunikative Validierung war aus organisatorischen Gründen nicht möglich.

Erwartungskonformität: Alle Personen dieses Untersuchungssamples beherrschten die grundsätzlichen Funktionen. Allerdings lässt sich daraus, aufgrund der kleinen Stichprobe, nicht ableiten, dass dies auch auf die Grundgesamtheit zutrifft. Die Computerkompetenz ist in weiteren Untersuchungen zu beobachten.

Perspektivübernahme: Die Instruktionen scheinen laut Interviewaussagen für die meisten Personen verständlich zu sein. Allerdings zeigen die Videos, dass diese Aussage nicht auf alle Probanden und Probandinnen zutrifft. Zwei Videos geben eindeutig Hinweise darauf, dass die Instruktion zwar angehört aber nicht entsprechend ausgeführt wurde. So wurde die Instruktionen „Klicken Sie die richtige Antwort an und tragen Sie den Buchstaben in das freie Feld“ bei der Aufgabe „Buchstabensuche“ nur teilweise beachtet. Bezüglich dieser Aufgabe lässt sich aus den Interviews und den Videos nicht ableiten, worin die Varianz des Verständnisses begründet ist, was zu folgenden Vermutungen führt:

Vermutung 1: Die Instruktionstexte sind zu lang und zu unverständlich.

Vermutung 2: Die Testsituation beeinflusst die Konzentrationsfähigkeit. Die Präsenz der betreuenden Person führt zu einer Nervositätssteigerung und zur Beeinträchtigung der Konzentration geführt haben. Die beeinträchtigte Konzentration beeinflusst wiederum das inhaltliche Verständnis und die Speicherung der Informationen.

In der Aufgabe „Simones Bestellung“ trug ein Proband keine Ziffern in die dafür vorgesehenen Felder ein. Die Felder für die Ziffern waren – im Vergleich zu den blau hinterlegten Feldern für die Buchstaben – grau hinterlegt. So ist zu

vermuten, dass die Felder für die Zahlen nicht als solche erkannt wurden, in denen etwas einzutragen ist.

Eindeutige Ursachen ließen sich durch weitere Studien mit der *Thinking Aloud*-Methode feststellen, indem Rückfragen zu den Aktionen der Probanden und Probandinnen während der Testdurchführung gestellt werden.

Erlernbarkeit: Das System scheint die Erlernung desselben zu unterstützen. Darauf deutet hin, dass die Probanden und Probandinnen überwiegend aussagten, sich im Laufe der Testung besser orientieren zu können. Auch zeigen die Videos, dass die Probanden und Probandinnen zum Ende der Testung bzw. nachdem sie ein paar Minuten in der Testumgebung interagierten, deutlich weniger Unterstützung einforderten. Es ist zu vermuten, dass dieser Lerneffekt noch stärker ausfällt, sobald alle Hilfefunktionen implementiert sind. In den Hilfefunktionen werden der Aufbau und die Funktionen jeweils itemformatspezifisch und seitenabhängig erläutert.

Joy of Use: Der Mehrheit der Teilnehmenden hat das Programm gut oder sehr gut gefallen. Auch das Storyboard wurde überwiegend positiv bewertet. Als Veränderungsvorschläge wurden der Wunsch nach mehr Videos, mehr Geschichten und kürzeren Texten genannt. Begründet sind diese Aussagen vermutlich darin, dass zum Zeitpunkt der ersten Usability-Studie (des zweiten Zyklus) die Testumgebung wenige Aspekte des Storyboards und nur zwei Videos beinhaltete. Die insgesamt äußerst positiven Äußerungen können zudem einerseits darin begründet sein, dass die Online-Testumgebung otu.lear mit dem Programm ich-will-lernen.de⁷¹ verglichen wird und dieses Programm den einzigen Referenzpunkt darstellt. Andererseits können diese positiven Rückmeldungen auch in der sozialen Erwünschtheit begründet sein. Auch wenn mehrfach betont wurde, dass nicht die funktionalen Analphabeten und Analphabetinnen selbst geprüft werden, wurde die Situation von ihnen mit großer Wahrscheinlichkeit als Kompetenztest empfunden. Folglich können sie das Gefühl gehabt haben, es gäbe eine richtige und eine falsche Antwort, wobei die richtige Antwort als diejenige empfunden wurde, die das Programm positiv (im Sinne der sozialen Erwünschtheit) bewertet. Weiterhin kann die direkte persönliche Kommunikation eine Hemmschwelle dahingehend sein, sich negativ über ein Produkt zu äußern, zu deren Entwickler- und Entwicklerinnen-Team ggf. die interviewende Person zählt. Des Weiteren wurden tendenziell geschlossene Fragen und teilweise Suggestivfragen gestellt. Die Frage „Wie gut fanden Sie das Programm“ lässt z. B.

⁷¹ Das Programm ich-will-lernen.de ist bei Teilnehmenden von Alphabetisierungskursen bekannt. Mit mehr als 31.000 Aufgaben (DVV) ist das Programm vergleichsweise umfangreich. Allerdings erscheint es wenig übersichtlich und insbesondere die Audiodateien weisen eine schlechte Qualität auf, so dass diktierter Wörter nicht verstanden und demnach von Lernenden nur durch Antizipation korrekt getippt werden können (vgl. zu „ich-will-lernen.de“ auch Kapitel 2.3.6).

nur eine Ausprägung der Bewertung „gut“ zu. Damit die Frage suggeriert, das Programm sei grundsätzlich gut, von den interviewten Personen wird nur die Antwort über das „wie gut“ gefordert. Auch das kann dazu geführt haben, dass die Personen sich tendenziell positiv geäußert haben. Für Folgeerhebungen ist zu beachten, Fragen offen zu formulieren, um differenziertere und aussagekräftigere Aussagen zu erhalten. Zudem sind die Interviewer und Interviewerinnen zu schulen, um beispielsweise Suggestivfragen zu vermeiden.

Die erste Forschungsfrage, die mit Hilfe dieser Studie beantwortet werden sollte (Welche und wie viel Unterstützung benötigen die Teilnehmenden bei der Bearbeitung der Aufgaben?), ist folgendermaßen zu beantworten: Die Erlernbarkeit von otu.lea ist hoch und das Programm ist in den Bereichen der Aufgabebearbeitung selbstbeschreibungsfähig. Jedoch weist der Bereich, in dem sich die Teilnehmenden anmelden sollen, eine niedrige Selbstbeschreibungsfähigkeit auf und muss dahingehen überarbeitet werden.

Die zweite Frage nach der Zufriedenheit der Teilnehmenden, ist mit „überwiegend zufrieden“ zu beantworten. Einige wenige Veränderungsideen und kritische Äußerungen werden in der weiteren Entwicklung berücksichtigt.

7.4 Zyklus 2: Re-Design

Auf Grundlage der Ergebnisse wurde die Testumgebung wie folgt modifiziert:

Heuristik	Modifizierungsvorschläge für otu.lea.
Selbstbeschreibungsfähigkeit	Der Instruktionstext bei der Codeeingabe wurde überarbeitet. Zudem wurde ein Screencast eingebunden, in dem die Codeeingabe erläutert wird. Das Screencast soll unter der Hilfefunktion – symbolisiert als Fragezeichen – abrufbar sein.
Erwartungskonformität	Die Computerfähigkeit scheint den Anforderungen gerecht zu werden. Für aussagekräftige Ergebnisse ist dies mit einer größeren Stichprobe zu überprüfen.
Perspektivübernahme	Die Instruktionstexte der Aufgaben „Buchstabensuche“ und „Simones Bestellung“ wurden im Sinne einer <i>accomodation</i> gekürzt.
Erlernbarkeit	Um die Erlernbarkeit des Systems zu steigern, wurde in einem Tutorial stärker auf die Hilfefunktion hingewiesen. Zudem wurden itemformatspezifische Hilfevideos eingebunden.
Joy of Use	Auch wenn sich nur eine Person negativ zu den Bildern der Charaktere geäußert hat, wurden diese etwas verkleinert. Dies ist auch darin zu begründen, dass sie teilweise die grafische Begrenzung der Testumgebung überragten. Zudem werden bei umfangreichen Texten, die im Storyboard begründet sind, Prescreens eingebunden. Diese werden vor der Aufgaben-seite angezeigt und präsentieren Informationen zum Storyboard.

Tabelle 17: Modifikationen von otu.lea auf der Grundlage des zweiten Zyklus

Ziel der Modifizierung bezüglich der Anmeldung (Login-Seite bzw. Codeeingabe) war eine selbstbeschreibungsfähige Einstiegssequenz. Vorerst wurde die Veränderung des Instruktionstextes fokussiert. Ein möglicher Einfluss der Computerkompetenz auf das Testergebnis konnte nicht festgestellt werden. Daher blieb die Vorannahme bestehen, dass keine weiteren Unterstützungs- oder auch Übungssequenzen angeboten werden müssen. Die Instruktionstexte wurden auf ihr Kürzungspotenzial hin überprüft. Die Texte, die nicht unmittelbar aufgabenbezogen sind und das Storyboard betreffen, wurden soweit möglich auf einem Prescreen (eine vor der eigentlichen Aufgabe platzierte Seite) dargestellt. Soweit eine Kürzung ohne Veränderung des Inhalts oder Reduzierung der Verständlichkeit möglich war, wurde diese umgesetzt. Es wurden für alle Interaktionsmöglichkeiten Videos entwickelt, in denen die Bearbeitung der Aufgaben und Bedeutung der Schaltflächen erläutert werden. Die kontextsensitiven Videos sollen auf den entsprechenden Seiten über die Hilfefunktion abrufbar sein. Die Grafi-

ken der Charaktere des Storyboards wurden verkleinert. Diese zuletzt genannte Modifikation ist neben der Interviewaussage (s. Tabelle 16, Proband/Probandin A2) darin begründet, dass die Bilder über den grafischen Seitenrand der Testumgebung hinausragten und somit einen Großteil des Interfaces einnahmen. Nach den genannten Modifizierungen wurde die Testumgebung weiteren laufende Evaluationen unterzogen. Eine stärkere Fokussierung der potenziellen Nutzer und Nutzerinnen wurde im dritten Zyklus vorgenommen.

Das Entwicklungsergebnis von otu.lea, welches auf Basis der Zyklen 1 und 2 sowie weiteren laufenden Evaluationen umgesetzt wurde, wird im Folgenden veranschaulicht:



Abbildung 20: Einstiegsseite mit Videotutorial von otu.lea

Aus der Perspektive der Nutzer und Nutzerinnen beginnt der Einstieg mit einem Videotutorial.

In einem halbminütigen Videotutorial werden den Nutzern und Nutzerinnen das Ziel und die Funktionen von otu.lea erläutert. Es wird beschrieben, welche Aktivitäten die Personen dort durchführen können, wie die Seiten aufgebaut sind und wo es Unterstützungsfunktionen gibt. Zudem werden der Ablauf und die Schaltflächen erklärt.

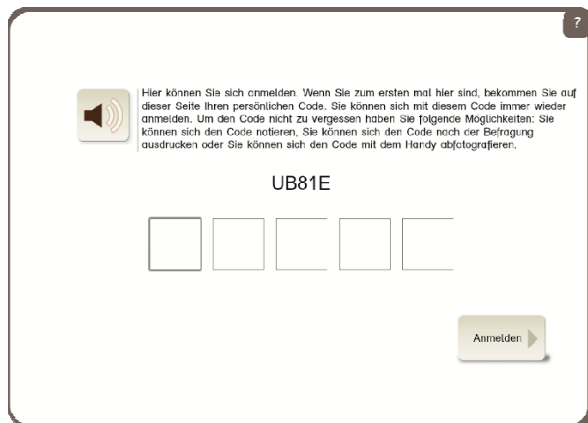


Abbildung 21: Anmeldung und Registrierung

Nach dem Einführungsvideo können sich die Teilnehmenden mit einem anonymen Code registrieren. Mit diesem können sie sich beliebig oft anmelden. Mit Hilfe des Codes kann der Lernprozess abgebildet werden, da dieser eine eindeutige Zuordnung zu den Testergebnissen einer Person und den Vergleich der Performanz mehrfacher Durchläufe ermöglicht. Der Erläuterungstext wird den Teilnehmenden nach Aktivierung der Schaltfläche vorgelesen. Um die Teilnehmenden bei der Dokumentation des Codes zu unterstützen wird ihnen der Hinweis gegeben, dass sie sich den Code merken, notieren oder abfotografieren können, um sich später mit dem gleichen Code anzumelden.



Abbildung 22: Dimensionsauswahl

Nach der Registrierung (entweder mit einem bereits vorhandenen Code oder der Neuanmeldung mit einem neuen Code) können die Teilnehmenden eine Dimension wählen.

Hier sind sogenannte Mouse Over integriert: Alleine durch das Navigieren mit dem Mauszeiger über die Schaltflächen wird eine Audiodatei abgespielt, welche den Titel der entsprechenden Dimension nennt.

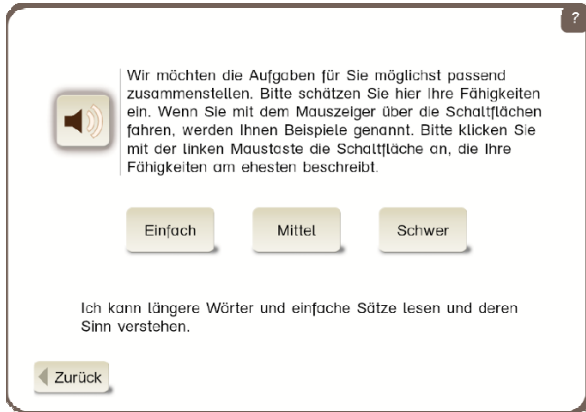


Abbildung 23: Selbsteinschätzung

Haben sich die Teilnehmenden für eine Dimension entschieden, nehmen sie eine Selbsteinschätzung hinsichtlich ihrer Kompetenzen in der entsprechenden Dimension vor.

Auch hier ist ein Mouse Over implementiert: Fahren die Nutzer und Nutzerinnen über eine Schaltfläche werden ihnen Beispiele für das Niveau vorgelesen. Navigiert eine Person beispielsweise über die Schaltfläche „Mittel“ wird ihnen Folgendes vorgelesen: „Ich kann längere Wörter und einfache Sätze lesen und deren Sinn verstehen.“

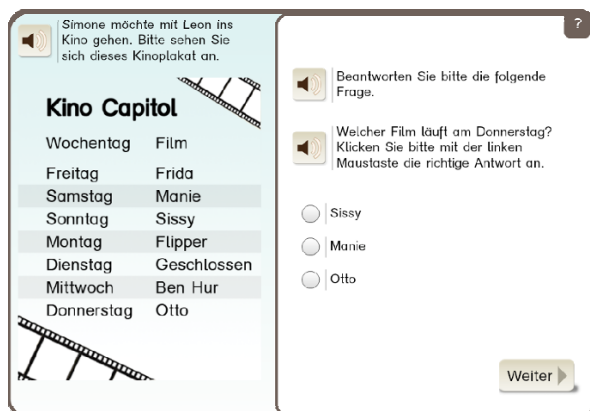


Abbildung 24: Itembeispiel aus der Dimension Lesen

Dem ausgewählten Niveau entsprechend wird das Aufgabenset ausgewählt und die Diagnostik beginnt. Das Interface für die Aufgabenbearbeitung ist immer gleich aufgebaut: Der linke Bereich dient der Darstellung von Einführungstexten und Stimuli in Form von Bildern. Auf der rechten Seite befindet sich der Aufgabenbereich.

Die Aufgaben bestehen in den meisten Fällen aus mehreren Items. Besonders in der Dimension Schreiben sind die Items durch das Storyboard miteinander verbunden. Beginnt eine Aufgabe mit einer näheren Erläuterung der Charaktere und deren Hintergrund, ist ein Prescreen vor der Aufgabenseite vorgeschaltet, auf dem lediglich Informationen abgebildet werden, die das Storyboard betreffen.



Abbildung 25: Kontextsensitive Hilfefunktion

Auf jeder Seite besteht die Möglichkeit, die Hilfefunktion abzurufen. Diese ist zum größten Teil auf den jeweiligen Kontext (die Seite, auf der sich ein Nutzer/eine Nutzerin momentan befindet) abgestimmt: Es werden entsprechend der Seite die Struktur, die Symbole und die Funktionen erläutert. Das Interface wird dabei ausgegraut, um einerseits die Aufmerksamkeit auf das Video zu lenken und andererseits zu ermöglichen, parallel zum Hilfevideo das Gesehene auf das momentane Interface zu übertragen.



Abbildung 26: Ende der Diagnostik

Am Ende eines Durchlaufs können Teilnehmende entscheiden, ob sie einen weiteren Test absolvieren oder die Umgebung verlassen möchten. Entscheiden sie sich dafür einen weiteren Durchlauf zu absolvieren, werden sie auf die Seite der Dimensionsauswahl geleitet und können erneut Dimension und Niveau wählen.

In der gesamten Umgebung sind visuelle und auditive Unterstützungen (Cues) in Form von blinkenden Schaltflächen und auditiven Erläuterungen integriert, um die Nutzer-/Nutzerinnenführung zu unterstützen und die Orientierung zu erleichtern. Auf jeder Seite ist eine Schaltfläche mit einem Fragezeichen eingebettet. Wird dieses aktiviert, wird ein Hilfevideo gezeigt, in dem in den meisten Fällen der dem Kontext entsprechende Seitenaufbau sowie die Funktionen (je nachdem, in welchem Teilbereich sich der Nutzer/die Nutzerin befindet) erläutert werden. Zum Zeitpunkt nach dem zweiten Evaluationszyklus waren allerdings noch nicht alle kontextsensitiven Hilfevideos entwickelt und eingebunden. Alle Audiodateien und Videos können beliebig oft abgespielt werden. Im Diagnosebereich ist allerdings die Möglichkeit ausgeschlossen, eine oder mehrere

Seiten zurückzugehen. Damit soll verhindert werden, dass sich ein möglicher Lerneffekt auf die Diagnostik auswirkt, indem z. B. eine Person auf späteren Seiten ein Wort liest, welches auf vorherigen Seiten geprüft wurde. Wäre das Zurückgehen möglich, könnte die Person den eigenen Eintrag auf einer vorherigen Seite korrigieren.

Die Rückmeldefunktionen waren zwar nach den ersten beiden Zyklen konzeptionell erstellt, jedoch konnten sie für den dritten Zyklus, die Usability-Studie 2013 (s. folgendes Kap.) noch nicht implementiert werden.

7.5 Reflexion der Usability-Studie 2012

Die zweite Vorstudie lieferte einerseits hilfreiche Ergebnisse, um die Testumgebung entsprechend der Bedürfnisse potenzieller Teilnehmer und Teilnehmerinnen zu gestalten. Insbesondere waren für die weitere Entwicklung von otu.lea die Login-Seite sowie das Konzept und der Inhalt für die Hilfefunktionen zu modifizieren. Andererseits konnten hilfreiche Hinweise bezüglich des Methodeneinsatzes gewonnen werden.

Bei der Auswertung der Videos wurde deutlich, dass die Betreuer und Betreuerinnen mitunter sehr schnell unterstützt haben. In einigen wenigen Fällen wurden das Interface und die nächsten vorzunehmenden Schritte erläutert, bevor die teilnehmende Person eigenständig aktiv werden konnte. Die Videos, aus denen das hervorgeht, konnten in der Auswertung somit nur bedingt berücksichtigt werden. Insbesondere sind diese Videos nicht auswertbar hinsichtlich der Selbstbeschreibungsfähigkeit. Um aussagekräftigere Ergebnisse zu erzielen könnten beispielsweise die betreuenden Personen noch stärker darauf aufmerksam gemacht werden, nicht sofort einzugreifen sobald der Proband/die Probandin nicht weiter weiß. Eine Alternative zur Erhebungsmethode wäre ein klassischer Usability-Test in Laboratories (vgl. Kap. 2.7.5), bei dem die Teilnehmenden über den Verlauf der Programmnutzung auf sich alleine gestellt sind und keine Unterstützung erhalten. Aufgrund der Zielgruppe mit einer vermutlich hohen Misserfolgsbefürchtung und fehlender Erfahrung in Bezug auf die Computernutzung erscheint diese Methode jedoch das Risiko des vorzeitigen Abbruchs zu bergen.

Bezüglich der Interviews ist anzumerken, dass nur wenige Informationen über die Gründe der von den Probanden und Probandinnen vorgenommenen Bewertungen gewonnen werden konnten. Zudem wurden die Interviewfragen vermehrt suggestiv gestellt, beispielsweise ist der Frage „Wie gut hat Ihnen das Programm gefallen?“ die Bewertung „gut“ implizit. Auch wurden den Teilnehmenden die Fragen „in den Mund gelegt“, mit Aussagen von den Interviewern und Interviewerinnen wie „...also sehr gut.“ Es ist zu vermuten, dass die Suggestivfragen einen großen Einfluss auf das Antwortverhalten hatten.

Um Suggestivfragen zu vermeiden, ist die Durchführung von Interviewschulungen und Übungseinheiten mit den Interviewern und Interviewerinnen sinnvoll. Für das Erfassen von Gründen für die Bewertung und ein differenziertes Meinungsbild ist der Einsatz von stärker geöffneten Fragen im Interview hilfreich.

7.6 Zusammenfassung und Reflexion hinsichtlich des dritten Zyklus

Die folgende Tabelle zeigt eine Zusammenfassung der Ergebnisse beider Vorstudien:

	Phase 1: Gestaltung	Phase 2: Durchführung	Phase 3: Überprüfung	Phase 4: Re-Design
Zyklus 1: Paper Prototyping	Entwurf des Interfaces und erster Aufgaben	Testung Paper Prototyping und fünf funktionalen Analphabeten und Alphabetinnen	Cues zur Nutzer-/Nutzerinnenführung erscheinen sinnvoll, auf die Hilfefunktion sollte stärker hingewiesen werden	Entwicklung eines ersten technischen Prototyps von otu.leaf
Zyklus 2: Usability-Studie 2012	Entwicklung der digitalen Version mit einem Aufgabensample, einer rudimentären Hilfefunktion sowie multimedialen Cues	Usability-Testing und Durchführung von fokussierten Interviews mit elf funktionalen Analphabeten und Alphabetinnen	Auf die Hilfefunktion ist stärker hinzuweisen, die Texte werden teilweise als zu lang empfunden, die Login-Seite ist zu überarbeiten, die Computerkompetenz erscheint für die Durchführung des Tests ausreichend	Die Hilfefunktion wird formatspezifisch implementiert, die Instruktionstexte werden gekürzt soweit das keinen inhaltlichen Einfluss hat, es wird versucht, die Login-Seite verständlicher zu gestalten

Tabelle 18: Zusammenfassung der ersten zwei Evaluationszyklen

Die ersten beiden Zyklen hatten das Ziel, Erkenntnisse sowohl auf der inhaltlichen als auch auf der methodischen Ebene zu liefern: Inhaltlich wurden in den zwei Studien die Erlernbarkeit, Selbstbeschreibungsfähigkeit, Wahrnehmungssteuerung, Erwartungskonformität, Perspektivübernahme sowie Joy of Use fokussiert. Die Ergebnisse wurden in der weiteren Entwicklung berücksichtigt und es fanden weitere Experten-/Expertinnen-Evaluationen statt (Heuristische Evaluationen, Cognitive Walkthroughs, Consistency Inspections). Auf der methodischen Ebene konnte festgestellt werden, dass erstens das „Evaluationspersonal“ genauere Anweisung dahingehend benötigt, ab welchem Zeitpunkt wie zu unter-

stützen ist. Zweitens sind die Interviewer und Interviewerinnen umfassender zu schulen.

Schließlich sind die Vorstudien und deren Ergebnisse in dem Kontext dieser Forschungsfrage zu verorten. Die Vorstudien hatten nicht zum Ziel, Hinweise darauf zu geben, ob geläufige Heuristiken auch auf funktionale Analphabeten und Analphabetinnen zutreffen. Vielmehr sind die Vorstudien eine Vorbedingung für die Durchführbarkeit der Hauptuntersuchung. Die Überprüfung (und im Idealfall Erfüllung) der Heuristiken diente der Entwicklung eines funktionsfähigen Prototyps und als Grundlage für die Beantwortung der Frage, ob spezielle Heuristiken für funktionale Analphabeten und Analphabetinnen entwickelt werden können.

7.7 Weitere Entwicklungsergebnisse der Online-Testumgebung

Weitere Entwicklungsbestandteile von otu.lea sind die technische Umsetzung, die systemgenerierte Auswertung, die Berücksichtigung und Umsetzung von Merkmalen des Rich E-Assessments sowie förderdiagnostischer Aspekte, die nicht Gegenstand der Überprüfungen in den Vorstudien waren.

7.7.1 Die Technologieplattform von otu.lea

otu.lea wurde als offen zugängliches System entwickelt. Das lea.-Diagnostikum ist in all seinen Dimensionen und Niveaustufen sowie fast allen Aufgaben und Items realisiert worden. Lediglich wenige Items konnten aus technischen Gründen nicht realisiert werden. Darüber hinaus wurde das System so gestaltet, dass auch in Zukunft die Integration von neuen Aufgaben der verfügbaren Aufgabenformate ohne programmiertechnische Systemanforderungen möglich ist. Dafür wurde ein *Framework* als Technologiebasis entwickelt, auf dem die Online-Testumgebung otu.lea implementiert:

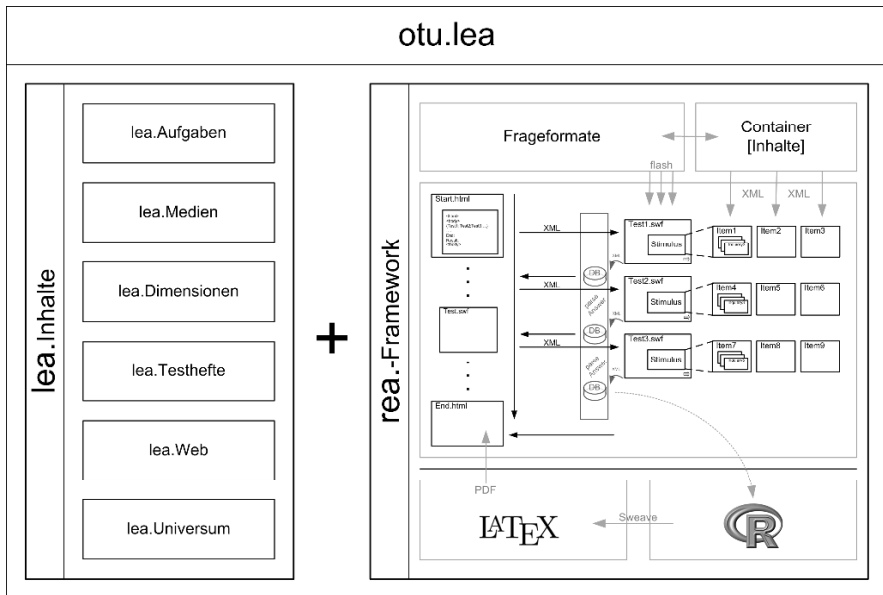


Abbildung 27: lea.-Framework (Wolf u. a., 2011, S. 136)

Grundvoraussetzung war die Verfügbarkeit von otu.lea über das Internet sowie die Rechner- und Browserunabhängigkeit. Das System läuft daher als Adobe Flash-Anwendung. Die Inhalte waren bereits in einer papierbasierten Form vorhanden und wurden an die technische Umgebung im Sinne einer *accomodation* angepasst.

Das Framework ist die generische Softwareimplementation des Diagnosewerkzeugs. Es besteht aus den Basiselementen Itemformat, Auswertungsalgorithmus und Inhalt. Die papierbasierten Itemformate wurden analysiert und als generelle Grundfunktion in der Software umgesetzt. Die Aufgabeninhalte wurden getrennt in strukturierten XML-Dateien codiert. Diese werden dann von der Software dargestellt („abgespielt“). Die Erweiterung oder Änderung der Inhalte hat damit keine Auswirkung auf die Funktion des Grundsystems. Hinzu kamen Anreicherungen in Form von Grafiken, Sound und Video. Die Auswertung wird mit der Statistiksoftware R vorgenommen und über L^AT_EX in eine lesbare Form übertragen und für die Ausgabe an funktionale Analphabeten und Analphabetinnen sowie Kursleitende als PDF generiert.

Das Framework wird nach Veröffentlichung der Plattform unter „open source“ als frei verfügbares und erweiterbares System zur Verfügung stehen.

7.7.2 *Systemgenerierte Auswertung mit otu.lea*⁷²

Durch die Kombination der Statistik-Software R (<http://www.r-project.org>) mit dem Programmpaket LaTeX (<http://www.latex-project.org>) und Sweave (<http://www.stat.uni-muenchen.de/~leisch/Sweave/>) sowie selbst entwickelter Auswertungsskripts werden aus der Online-Testumgebung Berichte erzeugt. In diesem wird der Kompetenzstand vor dem Hintergrund des Kompetenzmodells und auf der Grundlage der Diagnostik abgebildet. Möglich ist dies durch die Verknüpfungen jedes Items mit einer oder mehrerer Kannbeschreibungen und mit einem anonymisierten persönlichen Teilnehmenden-Code. Die Aufgaben werden dichotom ausgewertet, wodurch das System alle erfüllten, teilweise erfüllten und nicht erfüllten Kannbeschreibungen identifiziert und in einem Report für Teilnehmende und einem Report für Kursleitende zusammenträgt (das Entwicklungsergebnis der Feedbackfunktion wird in Kap. 8.5 näher vorgestellt). Die Berichte werden mit Textbausteinen über die statistische Auswertung und die dem Teilnehmenden zugeordneten XML-Antwortdateien erzeugt. Jede Rückmeldung kann als PDF-Datei generiert werden.

Die PDF-Datei wird dem Teilnehmenden direkt nach Abschluss der Diagnostik zur Verfügung gestellt. Sie enthält eine kurze Rückmeldung, eine Teilnahmebestätigung und den persönlichen Identifikations-Code. Der zweite Bericht ist eine detaillierte Auswertung, der für die Kursleitungen in den Bildungseinrichtungen bestimmt ist. Mit dem dahinterliegenden Kompetenzraster (Alpha-Level) kann eine individuelle Beratung und Planung der weiteren Förderung des Teilnehmenden durchgeführt werden. Anhand des Codes können einzelne Durchläufe miteinander verglichen werden und dazu dienen, den Lernprozess abzubilden. Für den Teilnehmenden entsteht damit ein Abbild seiner/ihrer Entwicklung. Für die Kursleiter und Kursleiterinnen in den Bildungseinrichtungen ist dies hilfreich bei der Ausarbeitung von Lernstrategien für ihre Kursteilnehmenden.

7.7.3 *Rich E-Assessment und Itemformate in otu.lea*

In Kap. 2.2.2 zum Thema Rich E-Assessment wurde bereits erläutert, dass das Befragungsmedium durchaus einen Einfluss auf die Validität haben kann. Einerseits stellt sich die Frage, ob durch augenscheinliche Veränderungen des Itemformats – vom Papier zum Computer – nicht doch Einfluss auf den Messgegenstand genommen wird. Andererseits ist fraglich, inwieweit die ICT-Literacy der Teilnehmenden ihr Antwortverhalten beeinflusst (vgl. auch Kap. 2.5 zu Äquivalenzproblematik).

⁷² Aspekte dieses Kapitels wurden bereits in Wolf, Koppel & Schwedes (2011) veröffentlicht.

Als Voraussetzungen und aus den Erfahrungen der ersten beiden Evaluationszyklen galt für die Entwicklung von otu.lean, eine angenommene niedrige ICT-Literacy der Teilnehmenden zu berücksichtigen und einer Monotonie entgegenzuwirken. Die Annahme einer niedrigen ICT-Literacy war zu treffen, um einerseits nicht durch komplexe Aufgabenformate, die komplexe Interaktion verlangen oder durch ein komplexes Interface-Design die Motivation negativ zu beeinflussen und womöglich einen Abbruch zu provozieren. Andererseits war darauf zu achten, dass eine niedrige Computerkompetenz nicht einen Einfluss auf das Antwortverhalten hat und somit die Validität beeinträchtigt. Den Herausforderungen (niedrige Literalität, möglichst hohe Abwechslung ohne Überforderung, niedrige ICT-Literacy und Motivationsförderung) wurde mit auditiven und visuellen Rich E-Assessment-Elementen begegnet.

Damit die Teilnehmenden die Instruktionen nicht selbst lesen müssen, wurden Audiodateien eingebunden. Mit den Audiodateien besteht kein Risiko, dass die Teilnehmenden die Aufgabe aufgrund ihrer Lesekompetenz nicht bearbeiten können, obwohl die Kompetenz zur Lösung der Aufgabe ggf. vorhanden ist. Die Nutzer und Nutzerinnen können konsistent gestaltete Schaltfläche in Form eines Lautsprechers anklicken und sich die Instruktion vorlesen lassen. Dieser Vorgang kann beliebig oft wiederholt werden, z. B. wenn Teilnehmende die Instruktion beim ersten Abspielen nicht verstanden haben oder auch, um nicht über einen längeren Zeitraum die Instruktionen im Arbeitsgedächtnis vorhalten zu müssen. Um einer Monotonie entgegen zu wirken und die Motivation zu steigern, wurde das Storyboard in die computerbasierte Version eingebunden. Das Storyboard lässt im Idealfall einen Spannungsbogen entstehen, der die Aufmerksamkeit und Durchhaltungsmotivation positiv beeinflusst. Zusätzlich existieren kontextsensitive Einführungs- sowie weitere Hilfevideos, die abgerufen werden können, wenn Unterstützung bezüglich der Handhabung des Programms notwendig ist (das Entwicklungsergebnis zur Hilfefunktion wird in Kap. 8.5 näher erläutert). Um Personen mit wenig Computererfahrung eine leichte Handhabung und Orientierung zu ermöglichen sind visuelle Cues in Form von blinkenden Schaltflächen eingebunden. Die blinkenden Schaltflächen sind nicht permanent, sondern in einer logischen Abfolge animiert. Zudem werden die Seiten iterativ aufgebaut: Um Aufgaben zu bearbeiten, werden zunächst nur der Lautsprecherbutton und ein Stimulus angezeigt. Der Lautsprecherbutton ist animiert (Cueing), um Aufmerksamkeit zu erzeugen. Sobald der Lautsprecherbutton aktiviert wurde, erscheint der Aufgabentext, der zugleich vorgelesen wird. Erst anschließend wird auf der rechten Seite des Interfaces die Aufgabe präsentiert – ebenfalls auditiv unterstützt. Die Schaltflächen und Aufgabensequenzen werden nacheinander animiert, z. B. blinkt zuerst der Lautsprecherbutton. Nach der Aktivierung des Lautsprecherbuttons und nachdem das Wort/der Satz vorgelesen wurde

blinkt das Textfeld, in welches das vorgelesene Wort/der vorgelesene Satz eingetragen werden muss. Im Anschluss an die Eingabe von Buchstaben wird die nächste Sequenz animiert, beispielsweise ein Feld zur Eingabe von Satzzeichen. Nachdem die Aufgaben bearbeitet wurden, blinkt der „Weiter-Button“, um zur nächsten Aufgabe zu gelangen.

Das Paper Prototyping-Experiment lieferte zwar Hinweise darauf, welche die Wirkung dieser Cues nahe legen (Wolf u. a., 2011). Allerdings ist weiter zu untersuchen, ob die Cues auch einen statistisch nachweisbaren Effekt erzielen (vgl. Evaluations-Zyklus 3).

Die Itemformate von otu.lea können anhand der in Kap. 2.2.6.2 vorgestellten Taxonomie analysiert und klassifiziert werden⁷³. Die Klassifizierung diene der systematischen Auswahl und Gestaltung der Itemformate unter Berücksichtigung des zu messenden Konstrukts und der Zielgruppe. In den Dimensionen Lesen, Sprachempfinden und Mathematik werden ausschließlich geschlossene Single Choice-Formate verwendet. Für die Dimension Schreiben werden geschlossene und halboffene Formate (Substitution/Correction und Completion) eingesetzt. Die Tabelle 19 stellt eine Übersicht der verwendeten Itemformate in otu.lea und der Ausprägungen der jeweiligen fünf systematisierenden Taxonomie-Dimension dar.

⁷³ Die Taxonomie baut auf der Taxonomie von Wolf, Koppel & Schwedes (2011) auf. Die Bewertungen der Formate haben sich aufgrund der Weiterentwicklung von otu.lea und der Erweiterung der Taxonomie leicht verändert.

Format Dimension	Single/ Multiple Choice	Selection/ Identifica- tion	Substitution/ Correction	Completion
1. Offenheit des	geschlossen	geschlossen	halboffen	halboffen
2. Auswertungsal- gorithmus				
Automatisierung	automatisiert	automatisiert	automatisiert	automatisiert
Komplexität	dichotom	dichotom	dichotom	dichotom
3. Authentizität	abstrakt bis teilkontextuali- siert	teilkontextu- alisiert	teilkontextuali- siert	teilkontextu- alisiert
4. Medienanreiche- rung				
Inhaltsdarstellung	mono- bis duocodal	duocodal	duocodal	duocodal
Medieneinbindung	wenig Varianz	wenig Vari- anz	wenig Varianz	wenig Vari- anz
Personalisierung	teilweise mög- lich	teilweise möglich	teilweise mög- lich	teilweise möglich
5. Interaktivität	mittel	mittel	mittel	mittel

Tabelle 19: Kategorisierung der *otu.lea*-Items in Anlehnung an Wolf, Koppel & Schwedes (2011, S. 138)

In *otu.lea* wurden Single/Multiple Choice-, Selection-/Identification-, Substitution-/Correction- sowie Completion-Formate eingesetzt. Single und Multiple Choice-Formate wurden in allen Dimensionen verwendet. In Lesen, Mathematik und Sprachempfinden wurden sie ausschließlich eingesetzt. Selection-/Identification-Formate werden selten in der Dimension Schreiben genutzt. Vorwiegend werden in der Dimension Schreiben Substitution- bzw. Correction- und Completion-Formate in Form von Textfeldern und Lückentexten eingesetzt.

1. *Offenheit des Antwortformats*: Die Single/Multiple Choice-Formate sind als geschlossen, Aufgaben, in denen etwas auszufüllen ist (Completion-Formate) zu klassifizieren. Aufgabenformate mit einer stark ausgeprägten Offenheit (wie z. B. freies Schreiben) werden bei *otu.lea* nicht eingesetzt.

2. *Auswertungsalgorithmus*: Der Auswertungsalgorithmus ist für alle Formate automatisiert umgesetzt. Jegliche Formate werden dichotom (richtig/falsch) ausgewertet und es liegt damit kein komplexer Auswertungsalgorithmus zugrunde.

3. *Authentizität*: Die Authentizität ist dimensionsabhängig unterschiedlich ausgeprägt. Für die Dimensionen Lesen und Schreiben sind die Formate durch den Bezug zu der Lebenswelt funktionaler Analphabeten und Analphabetinnen hinsichtlich der Charakteristiken „Art der Aufgabe“ und „Bewertungskriterien“ (vgl. Kap. 2.2.6.1 und Gulikers u. a., 2008) authentisch, die weiteren Charakteristiken („physikalischer Kontext“, „sozialer Kontext“ und „Ergebnispräsentation“) sind hingegen wenig bis gar nicht ausgeprägt. Daher sind die Items der Dimensionen Lesen und Schreiben als teilkontextualisiert einzustufen. In der Dimension Mathematik sind in den Aufgaben zwar Aspekte des Storyboards eingebunden, d. h. die Charaktere sind in einigen Aufgaben repräsentiert, jedoch sind die einzelnen Aufgaben nur wenig bzw. vorwiegend nicht durch einen Erzählstrang miteinander verbunden. Somit ist die Einbettung der Lebenswelt potenzieller Nutzer und Nutzerinnen und daher die Authentizität nur sehr gering und als abstrakt einzustufen. In der Dimension Sprachempfinden sind keine Anteile des Storyboards vorhanden. Auch sind in den Aufgaben kaum Anteile möglicher Lebensweltaspekte der potenziellen Nutzer und Nutzerinnen eingebunden. Daher sind die Aufgaben der Dimension Sprachempfinden ebenfalls als abstrakt zu bewerten.

4. *Mediyanreicherung*: Die Mediyanreicherung ist durch die Einbindung von Bildern, Audio, Animationen und Videos realisiert. Auch hier ist die Ausprägung der Mediyanreicherung in den einzelnen Dimensionen unterschiedlich ausgeprägt.

Die Single Choice Formate sind in Abhängigkeit der Dimension als mono- oder duocodal zu bewerten, da in den Dimensionen Lesen, Schreiben und Mathematik Text und Bilder verwendet werden, in der Dimension Sprachempfinden hingegen nur Text.

Für alle Formate und Dimensionen ist die Medienvarianz durch das Cueing, die Hilfefunktion mit kontextsensitiven Hilfevideos sowie die auditive Unterstützung vorhanden. Allerdings sind keine Videos oder Simulationen eingebunden, daher ist die Ausprägung der Medienvarianz als mittelmäßig und nicht als hoch einzustufen. Die Personalisierung ist für jegliche Dimensionen hinsichtlich der Lautstärkeneinstellung, der Niveau- und Dimensionswahl, der Unterbrechungsmöglichkeit sowie der Abrufbarkeit der Ergebnisse möglich. Allerdings besteht keine Personalisierungsmöglichkeit in Bezug auf die thematische Auswahl der Aufgaben oder der Erstellung eines personalisierten Nutzer-/Nutzerinnenprofils.

5. *Interaktivität*: Die Interaktivität ist bei den Formaten von otu.lea hinsichtlich der Schnelligkeit der Reaktion auf eine Aktion (durch das Anklicken von Schaltflächen und der darauf folgenden Reaktion seitens des Systems) gegeben. Aufgrund der komplexen Auswertung und der Möglichkeit, den Lernverlauf abbilden zu können, ist die Testumgebung interaktiv (vgl. Kap. 2.2.6.1 und Crawford, 2004). Die Wahlmöglichkeiten sind jedoch auf die Niveau- und Dimensionswahl begrenzt. Insgesamt ist die Interaktivität mäßig ausgeprägt.

Die Taxonomie diene zudem, passende Items für die Berücksichtigung förderdiagnostischer Aspekte zu entwickeln. Im folgenden Kapitel wird konkretisiert, wie förderdiagnostische Prinzipien in der Konzeption von otu.lea berücksichtigt wurden. Berücksichtigung förderdiagnostischer Aspekte bei otu.lea otu.lea stellt einen unverzichtbaren Teil des lea.-Förderdiagnoseinstruments – nämlich die Diagnostik – dar. Wie bereits im Kapitel zu Förderdiagnostik und zur computerbasierten Diagnostik erläutert, bietet eine computerbasierte Diagnostik das Potenzial förderdiagnostische Aspekte zu berücksichtigen. Im Folgenden werden die förderdiagnostischen Prinzipien, die in Kapitel 2.1.2 erläutert wurden, für die Reflexion der Online-Testumgebung hinsichtlich der Erfüllung der förderdiagnostischen Prinzipien herangezogen.

Die *Individualität* und die *Prozessorientierung* werden durch die Einbindung des Kompetenzmodells berücksichtigt: Auf der Basis der Alpha-Level und den Kannbeschreibungen geben die Ergebnisse Aufschluss über das individuelle Niveau. Es können dadurch sowohl weitere Tests anschließen als auch individuell passgenaue Fördermaßnahmen abgeleitet werden.

Den Prinzipien der *laufenden Feedbackprozesse* wird otu.lea gerecht, indem Rückmeldungen sowohl für Kursleitende als auch Teilnehmende automatisiert und unmittelbar erfolgen (vgl. Kap. 2.2.3 zu Vorteile einer computerbasierten Förderdiagnostik). Für die Kursleitenden wird der Entwicklungsprozess abgebildet, indem die Ergebnisse eines Nutzers/einer Nutzerin miteinander verglichen werden. Es wird angezeigt, ob sich eine Person im Vergleich zum vorherigen Durchlauf verbessert oder verschlechtert hat oder auf dem gleichen Niveau geblieben ist. Diese Darstellung ermöglicht die Unterstützung des förderdiagnostischen Prinzips *Prozessorientierung*. Für die Teilnehmenden wird ebenfalls eine Rückmeldung generiert, in der vermerkt ist, welche Kompetenzen bereits vorhanden sind und welche noch ausgebaut werden können. Die Rückmeldungen erfolgen sowohl schriftlich als auch auditiv. Mit der Berücksichtigung der Schwächen UND Stärken wird zudem ein weiteres Prinzip berücksichtigt (*Berücksichtigung von Stärken und Schwächen*).

Die laufenden Feedbackprozesse unterstützen ein weiteres förderdiagnostisches Prinzip: Das *Wechselverhältnis zwischen Diagnose und Intervention*. Die Diagnose kann zu beliebigen Zeitpunkten erfolgen, mit den Diagnostikergebnis-

sen können Interventionen zielgerichtet erfolgen, ggf. können Fördermaterialien angepasst oder Förderstrategien verändert werden. So ist eine genaue Passung zwischen Lernstandsdiagnose und Intervention möglich. Nur bedingt lässt sich das Prinzip Einbettung in das reale Umfeld berücksichtigen. Grundsätzlich lässt sich durch den Zugang über das Internet die Diagnostik in den „Alltag“ einbetten. Die Grafiken können allerdings nur exemplarisch das reale Umfeld der Teilnehmenden abbilden. Zudem ist das Storyboard vorgegeben und nicht auf individuelle Bedürfnisse modifizierbar.

Des Weiteren unterstützt die differenzierte Rückmeldung auf der Grundlage eines empirisch überprüften Kompetenzmodells in der Dimension Schreiben die *Durchführung von Fehleranalysen*. Durch die eingebundenen Theorie des Schriftspracherwerbs (Dluzak u. a., 2009; Frith, 1985) ist das Instrument theoretisch fundiert und kann zur Diagnose systematischer Fehler beitragen.

Werden die Überlegungen hinsichtlich der förderdiagnostischen Prinzipien in den internationalen Kontext eingebettet, ist im Sinne der förderdiagnostischen Attribute nach McManu (2008) kritisch anzumerken, dass ein Instrument nicht genuin Förderdiagnostik sein kann. Es bedarf der Einbettung in einen Beratungskontext, in dem die Kommunikation zwischen Lehrenden und Lernenden stattfindet (vgl. auch Bennett, 2011). Inhaltlich kann wiederum die differenzierte Diagnostik mit den Kannbeschreibungen zur Kommunikation beitragen, indem unter Berücksichtigung des Kompetenzstands Lernziele vereinbart werden.

Die förderdiagnostischen Prinzipien werden im Kontext dieser Studie nicht evaluiert. Möglich wäre dies, indem der Einsatz von otu.lea in Alphabetisierungskursen beobachtet und analysiert wird. Dabei sind ebenso die potenziellen Nutzer und Nutzerinnen als auch die Kursleitenden einzubeziehen. Im dritten Evaluationszyklus von otu.lea, der im Folgenden dargestellt wird, liegt der Fokus weiterhin auf der Usability von otu.lea.

8 Zyklus 3: Usability-Studie 2013

Der dritte Evaluationszyklus fand im Jahr 2013 statt. Zu diesem Zeitpunkt war otu.lea in großen Teilen funktionsfähig.

8.1 Zyklus 3: Gestaltung

Die Gestaltung und Konzeption des dritten Evaluationszyklus gliedert sich in drei Kapitel, in denen der Forschungsgegenstand, die Forschungsfragen sowie die Methodenwahl und die Bestimmung des Untersuchungskontextes detailliert erläutert werden.

8.1.1 *Entwicklung und Definition des Forschungsgegenstands*

Der dritte Zyklus setzt an dem Entwicklungsstand, der aus dem ersten und zweiten Zyklus (vgl. Kap. 6 und 7) sowie weiteren Modifikationen resultierte, an. In den dritten Zyklus sind damit die Ergebnisse der ersten beiden Zyklen eingeflossen. otu.lea befand sich zum Zeitpunkt des dritten Zyklus in einem Entwicklungsstadium, in dem die Items sowie die Umgebung weitestgehend implementiert und funktionsfähig waren (vgl. Beschreibung im vorherigen Kap. 7.7).

Gegenstand der Untersuchung des dritten Zyklus sind zentrale Elemente der Online-Diagnostik: Startseite, Niveauauswahl sowie ein Aufgabensample aus den Dimensionen Lesen und Schreiben. Bei der Auswahl der zu testenden Bereiche und Aufgaben wurde ein sinnvoller Kompromiss zwischen Breite (möglichst viele Aufgaben) und Tiefe (z. B. durch den Vergleich zwischen der Einbindung bzw. Nichteinbindung von Cues) gewählt. Kriterien für die Auswahl der Aufgaben waren Funktionalität und Aufgabenformat. Aufgaben, deren Funktionen nicht in vollem Umfang gewährleistet werden konnten, wurden ausgeschlossen. Die Hilfefunktion war in Form eines Videos implementiert und konnte während der Studie abgerufen werden. In dem Hilfevideo wurden die grundsätzlichen Funktionen zur Nutzung der Testumgebung sowie die Seitenstruktur erläutert. Die Hilfefunktion war bereichsübergreifend gestaltet, da vorerst weniger der inhaltlichen Effekt und vielmehr die generelle Nutzung der Hilfefunktion im Vordergrund stand⁷⁴. Die Feedbackfunktion war noch nicht in vollem Umfang funktionsfähig und wurde daher von der Evaluation ausgeschlossen.

⁷⁴ Die Umsetzung einer kontextsensitiven Hilfestellung wurde nach dem dritten Evaluationszyklus realisiert.

8.1.2 Formulierung der Forschungsfragen und Hypothesen

Der Entwicklungsstand von otu.lea war zum Zeitpunkt des dritten Evaluationszyklus so weit fortgeschritten, dass nun umfassende Forschungsfragen bearbeitet, weitere Forschungsmethoden eingesetzt und eine umfangreiche Studie durchgeführt werden konnten. Im Folgenden werden für den dritten Zyklus die Forschungsfragen, Hypothesen und deren Operationalisierung formuliert und differenziert dargestellt.

Die Forschungsleitfrage begründet sich zum einen in Desideraten aus der Praxis: Die Reduktion des funktionalen Analphabetismus mittels der Erhöhung der Teilnahmequote an Alphabetisierungskursen sowie die Effizienzsteigerung der Kurse selbst (für weitere Erläuterungen s. Kap. 2.3) und zum anderen aus mangelnden Erkenntnissen über Usability-Anforderungen für funktionale Analphabeten und Analphabetinnen (vgl. Kap. 2.6.4).

An den Herausforderungen in der Praxis anknüpfend ist die Forschungsleitfrage zum anderen aus theoretischen Zugängen abzuleiten: Um einen Beitrag zur Alphabetisierung zu leisten, den Herausforderungen in der Alphabetisierung zu begegnen (vgl. Kap. 2.3.4) und somit den funktionalen Analphabetismus zu reduzieren, kann eine Online-Diagnostik ein hilfreiches Instrument sein. Aufgrund einer tendenziell niedrigen ICT-Literacy von funktionalen Analphabeten und Analphabetinnen ist die Gebrauchstauglichkeit der Online-Diagnostik entsprechend dieser zielgruppenspezifischen Charakteristika zu gestalten. Dafür gilt es, sowohl Annahmen der CLT (vgl. Kap. 2.4) als auch der Usability-Forschung (vgl. Kap. 2.6) zu berücksichtigen.

Die zentrale Forschungsleitfrage lautet:

Wie muss eine Online-Diagnostik für funktionale Analphabeten und Analphabetinnen für den selbstständigen Gebrauch gestaltet sein?

Die Leitfrage differenziert sich in drei Bereiche und sieben Forschungsfragen:

1) Usability-Heuristiken

F1a: Werden die als für funktionale Analphabeten und Analphabetinnen relevant identifizierten Heuristiken in otu.lea berücksichtigt?

F1b: Gibt es Hinweise auf die Notwendigkeit, neue Heuristiken zu formulieren?

Um zu explorieren, was für funktionale Analphabeten und Analphabetinnen eine ausgeprägte Usability auszeichnet, werden die in Kap. 2.6.5 als relevant eingestuft Heuristiken als Evaluationsgrundlage herangezogen (F1a). Darüber hinaus wird das gesamte Datenmaterial des dritten Zyklus nach Auffälligkeiten analysiert, die darauf hindeuten, dass über die existierenden Heuristiken hinaus weitere Heuristiken für die Gewährleistung einer ausgeprägten Usability zu formulieren sind (F1b).

2) Computererfahrung und -kompetenz

F2a: Haben die Computererfahrung und -kompetenz einen Einfluss auf die Bedienbarkeit?

F2b: Haben die Computererfahrung und -kompetenz einen Einfluss auf die Performanz?

Falls F2b:

F2c: Was muss bei der Usability berücksichtigt werden, damit Computererfahrung und -kompetenz keinen Einfluss auf das Testergebnis haben?

Der zweite Fragenbereich wird mit Erhebungen zur Computererfahrung und -kompetenz, die mit weiteren Daten (Beobachtungen zur eingeforderten Unterstützung, zur Bearbeitungszeit (F2a) und Performanz (F2c)) in Beziehung gesetzt werden, bearbeitet.

Die Fragen begründen sich darin, dass für die Generierung aussagekräftiger Daten keine weiteren Faktoren einen Einfluss auf das Testergebnis haben dürfen, die nicht Bestandteil der Messung sein sollen. Der aktuelle Forschungsstand indiziert, dass die Computererfahrung und -kompetenz mögliche Einflussfaktoren bezüglich der Bedienung und der Performanz darstellen. Die Computererfahrung und Computerkompetenz sind daher als Kovariaten in den Untersuchungen zu berücksichtigen. Ziel ist über die Ergebnisse und Beobachtungen sicherzustellen, dass die Computererfahrung und -kompetenz keinen Einfluss auf die Performanz der Literalitätsmessung ausüben (vgl. auch Äquivalenzproblematik Kap. 2.5). Aus den vorangegangenen Überlegungen und den bisherigen Forschungserkenntnissen werden folgende Hypothesen aufgestellt:

H0_{a1}: Es besteht kein Zusammenhang zwischen den Computererfahrungen und der eingeforderten Unterstützung.

H1_{a1}: Es besteht ein Zusammenhang zwischen den Computererfahrungen und der eingeforderten Unterstützung: Personen mit wenig Computererfahrung werden öfter unterstützt als Personen mit viel Computererfahrung.

H0_{a2}: Es besteht kein Zusammenhang zwischen der Computererfahrung und der Aufgabenbearbeitungszeit.

H1_{a2}: Es besteht ein Zusammenhang zwischen den Computererfahrungen und der Aufgabenbearbeitungszeit: Personen mit wenig Computererfahrung benötigen mehr Zeit für die Aufgabenbearbeitung als Personen mit viel Computererfahrung.

H0_b: Es besteht kein Zusammenhang zwischen den Computererfahrungen und der Performanz.

H1_b: Es besteht ein Zusammenhang zwischen den Computererfahrungen und der Performanz: Je höher die Computererfahrung, desto höher die Performanz.

Sollte ein Zusammenhang festgestellt und die Nullhypothesen abgelehnt werden, ist zu hinterfragen, was berücksichtigt und ggf. eingebunden werden müsste, damit die Computererfahrung und -kompetenz keinen Einfluss auf die Bedienbarkeit und das Testergebnis ausüben⁷⁵.

3) Beanspruchung des *Extraneous Load*

F3a: Gibt es Hinweise auf die Beanspruchung des *Extraneous Load* durch den Einsatz von Cues?

F3b: Gibt es weitere Hinweise auf Faktoren, die den Cognitive Load im Allgemeinen und den *Extraneous Load* im Besonderen beeinflussen?

⁷⁵ Auch wenn die Thesen vordergründig suggerieren, dass ein Zusammenhang aufgedeckt werden soll, ist das Ziel, dass kein Zusammenhang festgestellt wird. Wie auch aus den vorherigen Überlegungen in Kapitel 2.8 hervorgeht, handelt es sich bei der Computererfahrung und -kompetenz um Kovariaten und somit um eine Art „Kontrollhypothesen“. Das interessierende Merkmal ist also „kein Zusammenhang“ bzw. „Zusammenhangslosigkeit“. Allerdings sind aus testtheoretischen Gründen die Alternativhypothesen (H₁) als Zusammenhangs- bzw. Unterschiedshypothesen zu formulieren (vgl. z. B. Bortz & Döring, 2006, S. 600; Kuckartz, Rädiker, Ebert & Schehl, 2013, S. 144 ff).

Ausgehend von Erkenntnissen der CLT ist anzunehmen, dass die kognitive Beanspruchung höher ist, wenn die Computererfahrung und -kompetenz niedrig sind (vgl. Kap. 2.3.5). Es gibt Hinweise darauf, dass multimediale Unterstützungsfunktionen die Beanspruchung des *Extraneous Load* reduzieren.

Die zu diesem Thema bereits existierenden Forschungsergebnisse werden auf die Forschungsfrage dieser Arbeit übertragen und es werden Zusammenhangshypothesen gebildet:

H0_{c1}: Es besteht kein Zusammenhang zwischen dem Einsatz von Cues und der Performanz.

H1_{c1}: Es besteht ein Zusammenhang zwischen dem Einsatz von Cues und der Performanz: Der Einsatz von Cues hat einen positiven Einfluss auf die Performanz.

H0_{c2}: Es besteht kein Zusammenhang zwischen dem Einsatz von Cues und der Aufgabenbearbeitungszeit.

H1_{c2}: Es besteht ein Zusammenhang zwischen dem Einsatz von Cues und der Aufgabenbearbeitungszeit: Der Einsatz von Cues führt zu einer geringeren Bearbeitungszeit.

Darüber hinaus ist anzunehmen, dass die Interviews und der UEQ weitere Hinweise zu potenziellen Auslastungsfaktoren des *Extraneous Load* im Besonderen sowie des *Cognitive Load* im Allgemeinen liefern.

Für die Bearbeitung der Fragen sowie Überprüfung der Hypothesen werden die Begriffe Usability, Computererfahrung und -kompetenz, Performanz sowie die Einsatzmöglichkeit(en) der Cues im folgenden Kapitel operationalisiert. In Erweiterung zur Darstellung der Wirkungszusammenhänge auf der theoretischen Basis (s. Abbildung 7 in Kap. 2.8) werden die Hypothesen folgendermaßen integriert:

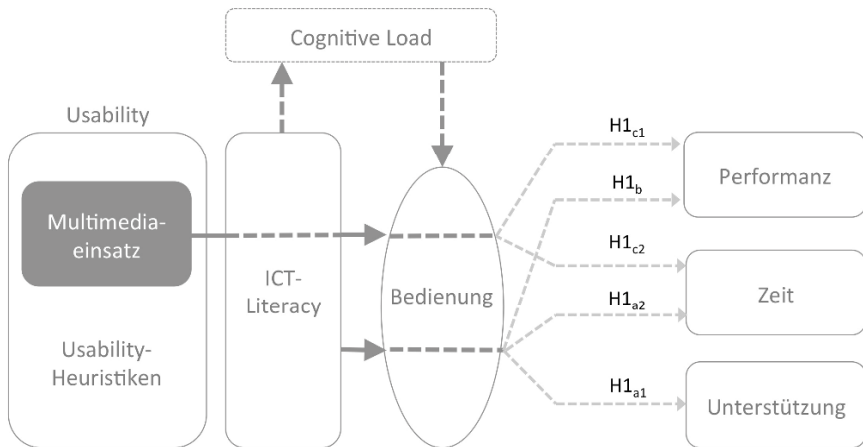


Abbildung 28: Wirkungszusammenhänge und Hypothesen

Die Abbildung verdeutlicht und konkretisiert die vermuteten Zusammenhänge. Über die Bedienung von otu.lea (mit Hilfe von Usability-Tests und deren Videographie) werden die Performanz, die Bearbeitungszeit sowie die Unterstützung beobachtet und auf Zusammenhänge hin untersucht. Über die hier konkret dargestellten Zusammenhänge hinaus wird die Online-Testumgebung auf der Grundlage der Heuristiken evaluiert.

8.1.2.1 Operationalisierung der Forschungsfragen und der Hypothesen

Die Operationalisierung der Forschungsfragen und Hypothesen erfolgt anhand der drei Fragebereiche Usability-Heuristiken, Computererfahrung und -kompetenz sowie *Extraneous Load*.

1) Usability-Heuristiken

F1a „Werden die als für funktionale Analphabeten und Analphabetinnen relevant identifizierten Heuristiken in otu.lea berücksichtigt?“: Für die Evaluation von otu.lea und die Operationalisierung der Forschungsfrage werden zehn der 15 vorläufigen Heuristiken für funktionale Analphabeten und Analphabetinnen als Evaluationsgrundlage herangezogen, die in Kap. 2.6.5 als zielgruppen- und themenrelevant identifiziert wurden. Fünf Heuristiken sind aufgrund von intern durchzuführenden Evaluationen und/oder Systemgrenzen auszuschließen: Die Heuristiken Konsistenz und Datensicherheit wurden intern mit Software-Experten/-Expertinnen evaluiert, da einerseits die Kriterien für die Konsistenz objektiv überprüfbar sind und andererseits das Thema Datensicherheit vorerst nicht auf Ebene der potenziellen Nutzer und Nutzerinnen zu evaluieren ist. As-

pekte zum Feedback konnten nicht überprüft werden, da die Feedbackfunktion zum Zeitpunkt der Studie noch nicht implementiert war. Für die Fehlertoleranz und Individualisierbarkeit sind seitens des Systems Grenzen vorhanden, die keine Modifikationen erlauben. Daher wurden diese Heuristiken nicht als Evaluationsgrundlage für die Studie herangezogen und sind nicht Bestandteil dieser Untersuchung.

Im Folgenden werden die studienrelevanten Heuristiken operationalisiert (für die Erläuterung der Heuristiken s. Kap. 2.6.4 und Kap. 2.6.5).

1. **Erlernbarkeit:** Für Erlernbarkeit ist anzeigend, ob eine Person anfangs mehr Hilfe (über die Hilfefunktion oder den Betreuer/die Betreuerin) eingefordert hat als gegen Ende des Testdurchlaufs. Würde gegen Ende die Unterstützungseinforderung ebenso hoch oder höher als zu Beginn sein, wäre das ein Indiz dafür, dass otu.lea einerseits nicht konsistent gestaltet ist bzw. die Probanden und Probandinnen vor neuen Herausforderungen stehen, weil Aspekte vorher nicht auftraten und/oder nicht erläutert wurden. Andererseits kann das darauf hindeuten, dass die Erläuterungen der Funktionen und die Hilfefunktionen nicht zum Verständnis beitragen. Beides wären Hinweise dafür, dass die Funktionen, die Struktur und die „Logik“ von otu.lea nicht oder nur schlecht gelernt werden können. Es wird vermutet, dass die Erlernbarkeit insbesondere bei Personen sichtbar wird, die keine oder wenig Computererfahrungen⁷⁶ haben. Ob stärker zu Beginn oder zum Ende Unterstützung eingefordert wird, wird mit Hilfe der Videos und der Gesamtbearbeitungszeit⁷⁷ erhoben.
2. **Selbstbeschreibungsfähigkeit:** Für die Beobachtungen und Beurteilung der Selbstbeschreibungsfähigkeit werden die Login-Sequenz und der Aufgabenbereich aufgrund der unterschiedlichen Funktionsweisen und Anforderungen getrennt analysiert. Für den Aufgabenbereich sind die Nutzung und der Effekt der Hilfefunktion anzeigend für die Selbstbeschreibungsfähigkeit. Es wird dokumentiert, wie oft eine Person die

⁷⁶ Zur Operationalisierung der Computererfahrung s. Operationalisierung der Forschungsfrage 2 im folgenden Abschnitt.

⁷⁷ Die Gesamtbearbeitungszeit unterscheidet sich von der aufgabenbezogenen Bearbeitungszeit: Die Log-Daten liefern lediglich die Bearbeitungszeit der Aufgaben. Die Zeit zwischen Einloggen und Bearbeitung der ersten Aufgabe ist nicht berücksichtigt. Zur Bearbeitungszeit wurde daher die Zeit ab dem Ende des Videotutorials bis zum Bearbeitungsbeginn der ersten Aufgabe hinzugerechnet. Der Zeitraum zwischen Beginn der Aufnahme bis zum Ende des Videotutorials muss somit von dem Zeitpunkt der Unterstützung abgezogen werden. Die Unterstützungen, die bereits während oder vor dem Videotutorial beobachtet wurden, werden nicht berücksichtigt, da sich die Unterstützung nicht direkt auf die Umgebung von otu.lea bezieht, sondern vorwiegend organisatorische und allgemeine Aspekte beinhaltet.

Hilfefunktion nutzt und ob die Hilfefunktion dazu beiträgt, otu.lea selbstständig weiter zu nutzen. Für die Login-Sequenz wird zudem berücksichtigt, wie oft Probanden und Probandinnen die Unterstützung von der betreuenden Person einfordern. Hierfür werden die Videos analysiert.

3. **Wahrnehmungssteuerung:** Die Wahrnehmungssteuerung wurde zum einen mit laufenden heuristischen Evaluationen überprüft. Zum anderen ist für die Wahrnehmungssteuerung anzeigend, ob der Einsatz von Cues einen Effekt auf die Aufgabenbearbeitungszeit hat.
4. **Aufgabenangemessenheit**⁷⁸: Die Aufgabenangemessenheit wird über die subjektiv empfundene Schwierigkeit der Aufgabenbearbeitung sowie über die wahrgenommenen Herausforderungen erhoben. Aussagen zur Verständlichkeit des Programms, insbesondere der Instruktionen sowie zu Problemen und Schwierigkeiten, die dazu führten, dass Aufgaben nicht bearbeitet wurden, sind ebenfalls anzeigend für die Angemessenheit der Aufgaben.
5. **Erwartungskonformität:** Hierfür ist ausschlaggebend, ob Personen auch mit niedriger Computererfahrung das Programm subjektiv als für sie nutzbar beurteilen. Hinweisend sind Aussagen von Probanden und Probandinnen, in denen sich zu Schwierigkeiten und Problemen geäußert wird, die sich nicht auf die inhaltliche Bearbeitung der Aufgaben beziehen.
6. **Steuerbarkeit:** Anzeigend für die Steuerbarkeit sind Aussagen über die fehlende Möglichkeit zur Unterbrechung, zur rigiden Führung (das Hin- und Herspringen zwischen Aufgaben ist nicht möglich) und darüber, ob sie die Ansagen als zu langsam/schnell empfanden.
7. **Perspektivübernahme:** Aussagen zur Verständlichkeit des Inhalts sollen Hinweise liefern, ob Voraussetzungen der potenziellen Nutzer und Nutzerinnen berücksichtigt wurden.
8. **Prozessangemessenheit:** Anzeigend für die Prozessangemessenheit sind Aussagen darüber, ob otu.lea für den Einsatz in Alphabetisierungskursen oder anonym, z. B. zu Hause, passend gestaltet ist. Hierfür werden Aussagen eines Experten-/Expertinnengesprächs mit Kursleitern und Kursleiterinnen und Aussagen von Probanden und Probandinnen herangezogen.
9. **Joy of Use:** Das Erleben während der Nutzung wird zum einen über Interviews und Aussagen über die Zufriedenheit erfragt. Zum anderen

⁷⁸ Unter Aufgabenangemessenheit ist sowohl zu verstehen, ob die Aufgabenschwierigkeit als angemessen empfunden wurde, als auch ob die Probanden und Probandinnen mit otu.lea das Ziel – nämlich die Bereiche Lesen und Schreiben bis zum Ende zu bearbeiten – erreichen konnten.

wird die Heuristik Joy of Use über Indikatoren zur Attraktivität bewertet, die mit dem UEQ erhoben werden.

10. **Interkulturelle Aspekte:** Ob interkulturelle Aspekte berücksichtigt werden, wird über Fragen zum Identifikationspotenzial der Teilnehmenden mit den lea.-Charakteren ermittelt.

Es sei erneut darauf hingewiesen, dass über die Überprüfung der Heuristiken hinaus danach gefragt wird, ob die Notwendigkeit für die Formulierung weiterer Heuristiken besteht, die nicht durch die evaluierten Heuristiken abgedeckt sind. Da die Heuristiken nicht speziell für funktionale Analphabeten und Analphabettinnen entwickelt wurden, können sie lediglich als eine Art Leitfaden für die Operationalisierung und schließlich der Formulierung zielgruppenspezifischer Heuristiken dienen. Es ist zu erwarten, dass über die auf der theoretischen Grundlage formulierten Heuristiken hinaus Aspekte für eine ausgeprägte Usability eine Rolle spielen.

F1b „Gibt es Hinweise auf die Notwendigkeit, neue Heuristiken zu formulieren?“: Das Datenmaterial des dritten Zyklus wird über die als auf theoretischer Basis relevant identifizierten Heuristiken hinaus hinsichtlich Auffälligkeiten gesichtet, die nicht einer der Heuristiken zugeordnet werden können. Als auffällig werden Aussagen und/oder Videosequenzen bezeichnet, die auf Unterstützungsnotwendigkeiten oder mangelnde Verständlichkeit der Online-Testumgebung hinweisen.

2) Computererfahrung und -kompetenz

Die Computererfahrung und -kompetenz werden über Befragungen zum Nutzungsverhalten und der Nutzungsfähigkeiten operationalisiert.

Zur Computererfahrung zählen die Nutzungsdauer (z. B. wie viel Zeit verbringt eine Person wöchentlich am Computer und im Internet) und -gewohnheiten (z. B. was für Aktionen führt die Person im Internet aus). Die Computerkompetenz wird hinsichtlich der geforderten Kompetenzen für die Bedienbarkeit von otu.lea geprüft. Anhand von drei Fragen werden die grundlegenden Computerkompetenzen, die für die Nutzung von otu.lea notwendig sind (Eingabe einer Internetadresse, Eingabe von Buchstaben in ein Textfeld sowie die Funktionsbedeutung und Bedienung der Hochstelltaste) erhoben (s. Kap. 1.1 im Anhang).

F2a „Haben die Computererfahrung und -kompetenz einen Einfluss auf die Bedienbarkeit?“: Die Bedienbarkeit wird anhand der eingeforderten Unterstützung und der Aufgabenbearbeitungszeit bewertet. Es wird geprüft, ob Personen mit niedriger Computererfahrung und/oder -kompetenz tendenziell mehr Unterstützung einfordern ($H1_{a1}$) und mehr Zeit benötigen als Personen mit hoher Computererfahrung und/oder -kompetenz ($H1_{a2}$). Zur Beantwortung dieser Frage wird ausschließlich die Aufgabenbearbeitungszeit herangezogen, da die Bearbei-

tungszeit der gesamten Testumgebung das Einführungstutorial sowie die Login-Sequenz einschließen würde. Diese beiden Sequenzen sind mehreren Einflussfaktoren ausgesetzt, wie beispielsweise Begrüßung, einführende Erläuterungen zur gesamten Studie oder Anschluss des Kopfhörers und Einstellung der Lautstärke. Daher ist die gesamte Bearbeitungszeit (aufgabenunabhängig) nicht aussagekräftig für die Beantwortung dieser Frage.

F2b „Haben die Computererfahrung und –kompetenz einen Einfluss auf die Performanz?“: Für die Überprüfung eines Zusammenhangs zwischen der Computererfahrung/-kompetenz und der Performanz ($H1_b$) dienen (neben den Ergebnissen zur Computererfahrung/-kompetenz) die Performanzwerte. Die Performanzwerten geben Aufschluss über die literalen Kompetenzen und werden über das lea.-Kompetenz-Modell erhoben. Das Kompetenzmodell der jeweiligen Dimension besteht aus einzelnen Kompetenzbeschreibungen – sog. Kannbeschreibungen. Die Dimension Schreiben wurde zudem anhand der probabilistischen Testtheorie empirisch validiert. Die Kalibrierung der Items erfolgte mit der papierbasierten Version auf der Grundlage des Rasch-Modells (vgl. z.B. Moosbrugger, 2008) und dient hier als Benchmark⁷⁹. Für die Items der Dimension Lesen wurde ein Schwierigkeitsindex entwickelt, anhand dessen die Items auf der Grundlage der Ergebnisse des Pre-Tests eingestuft wurden. Für Beantwortung der Frage werden die Items dichotom ausgewertet. Um eine Vergleichbarkeit über alle Probanden und Probandinnen und Dimensionen hinweg zu ermöglichen und die Bedingungen für inferenzstatistische Bewertungen zu gewährleisten, werden relative Performanzwerte gebildet. Hierfür wird ein Prozentwert für die gelösten Aufgaben ausgegeben.

F2c „Was muss bei der Usability berücksichtigt werden, damit die Computererfahrung und –kompetenz keinen Einfluss auf das Testergebnis haben?“: Falls bei der Beantwortung der Fragen F2a und/oder F2b ein Effekt festgestellt wird, werden die Ergebnisse sowie die Videos des dritten Zyklus hinsichtlich möglicher Ursachen analysiert und interpretiert: Es wird einerseits untersucht, ob bestimmte Aufgabenformate für einen Effekt verantwortlich sind und ob bei bestimmten Formaten ein tendenziell niedrigerer Performanzwerte erzielt wurde als bei weiteren eingesetzten Formaten. Andererseits werden die Videos hinsichtlich auffälliger Antwortverhalten analysiert. Auf Basis dieser Daten sollen zudem Faktoren identifiziert werden, die womöglich zu einer Reduktion dieser Effekte beitragen können.

⁷⁹ Ob die Kalibrierung mit dem computerbasierten Diagnoseinstrument das gleiche Ergebnis wie eine Kalibrierung mit dem papierbasierten Instrument erzeugt, ist nicht Bestandteil dieser Arbeit und wäre in weiteren Untersuchungen zu überprüfen.

3) Beanspruchung des *Extraneous Load*

F3a „Gibt es Hinweise auf die Verminderung der Beanspruchung des *Extraneous Load* durch den Einsatz von Cues?“: Der *Extraneous Load* wird unter Berücksichtigung des Einsatzes von Cues über die Aufgabenbearbeitungszeit, die Performanz, die Items des UEQs sowie die subjektiv wahrgenommene Schwierigkeit der Aufgaben und der Nutzung von *ota* operationalisiert. Um Hinweise auf die Beanspruchung des *Extraneous Load* zu gewinnen, werden für die Messung der Bearbeitungszeit und der Performanz mit Hilfe des Data-Logging objektive Messmethoden herangezogen. Für die Überprüfung der Hypothesen werden in die Datenerhebung experimentelle Bedingungen eingefügt: Cues an/Cues aus. Genauer werden die Bedingungen im Kap. 8.3.7 und Kap. 8.3.8 erläutert.

Um zu überprüfen, ob zwischen dem Einsatz von Cues und der Performanz ein Zusammenhang besteht ($H1_{c1}$), werden die relativen Performanzwerte für inferenzstatistische Berechnungen genutzt.

Ob ein Zusammenhang zwischen dem Einsatz von Cues und der Bearbeitungszeit besteht ($H0_{c2}$ bzw. $H1_{c2}$), wird über die aufgabenbezogene Bearbeitungszeit in inferenzstatistische Berechnungen einbezogen.

F3b „Gibt es weitere Hinweise auf Faktoren, die den Cognitive Load im Allgemeinen und den *Extraneous Load* im Besonderen beeinflussen?“: Darüber wird in weitere Überlegungen und der Interpretation der Ergebnisse das Vorwissen (in diesem Fall die Computerkompetenz und –erfahrung) berücksichtigt (vgl. Kap. 2.4 Theorieteil zu Cognitive Load). Hierfür wird geprüft, ob die Verteilung der Personen mit viel und wenig Computererfahrung hinsichtlich der Cues ungleich ist. Die Verteilung ist zu überprüfen, die Computererfahrung als Moderatorvariable auszuschließen ist⁸⁰.

Um Aufschluss über die wahrgenommene Schwierigkeit und unterstützende Faktoren zu erhalten, werden über die Interviews und über Ausschnitte des UEQ subjektive Messmethoden genutzt. Die Interviews werden im Hinblick auf Aussagen zur wahrgenommenen aufgaben- und nutzungsbezogenen Schwierigkeit analysiert. Aus dem UEQ werden aus der Skala „Durchschaubarkeit“ die Gegensatzpaare „verständlich – unverständlich“, „leicht zu lernen – schwer zu lernen“, „kompliziert – einfach“ und „übersichtlich – verwirrend“ herangezogen.

Die Untersuchungen fanden im Rahmen von Alphabetisierungskursen statt. Durch den institutionellen Untersuchungskontext (z. B. räumliche Gegebenheiten, Computerausstattung) und die Teilnehmenden (z. B. Testangst, niedrige

⁸⁰ Eine ungleiche Verteilung kann ein Hinweis darauf sein, dass die Ergebnisse durch die Computererfahrung moderiert werden und daher die Computererfahrung als Einflussfaktor in weitere Überlegungen einzubeziehen ist.

Literalität) ergaben sich Bedingungen, die in den methodischen Vorüberlegungen sowie der Methodenwahl zu berücksichtigen sind. Die konkreten Voraussetzungen und schließlich die Methoden- und Instrumentenwahl werden in den folgenden Kapiteln erläutert.

8.1.2.2 Methodische Vorüberlegungen

Hinsichtlich der Methodenwahl und dessen Einsatzform wurden die Rahmenbedingungen sowie die Charakteristika der Zielgruppe berücksichtigt. Der Zugang zur Zielgruppe ist mit einem vertretbaren Aufwand nur über Alphabetisierungskurse möglich, da funktionale Analphabeten und Analphabetinnen außerhalb eines institutionellen Kontextes nicht oder nur vereinzelt und mit hohem Recherche-Aufwand zu erreichen sind. Alphabetisierungskurse haben ein festes Zeitfenster, das in Absprache mit den Kursleitenden für die Studie zwar voll ausgeschöpft, nicht aber überschritten werden darf. Auch die zeitliche Streckung der Studie auf beispielsweise zwei Termine wurde abgelehnt, da sowohl mit einer sinkenden Akzeptanz als auch mit organisatorischen Schwierigkeiten zu rechnen wäre. Der zeitliche Umfang ist aufgrund dessen auf höchstens 1,5 Stunden begrenzt. Es ist davon auszugehen, dass die Probanden und Probandinnen hinsichtlich ihrer literalen Kompetenzen und Computererfahrungen heterogen sind, sie eine geringe Aufmerksamkeitsspanne und hohe Testangst haben. Daher werden diverse Kompetenzstufen bedient und das Risiko einer Überforderung minimiert. Die Befragungen werden in einfacher Wortwahl durchgeführt. Hinsichtlich des Untersuchungssettings wird auf eine angenehme und freundliche Atmosphäre geachtet, in der die Teilnehmenden bestärkt werden und nicht ihre literalen bzw. computerbezogenen Defizite im Vordergrund stehen.

Wie bereits aus vorherigem Kapitel hervorgeht, wird für die Beantwortung der Forschungsfragen sowohl explorativ (1) als auch hypothesengeleitet (2) vorgegangen. Entsprechend wird für die Gewinnung der Daten ein Mixed Method-Ansatz gewählt (vgl. z. B. Johnson u. a., 2007). Es geht hier nicht um die Erfassung desselben Merkmals mit unterschiedlichen Instrumenten, wie es die Triangulation beschreibt (Bortz & Döring, 2006, S. 365), sondern um die Ergänzung verschiedener Methoden, um eine umfassende Evaluation zu ermöglichen.

„[...] quantitative techniques [...] are helpful in identifying broad patterns, but qualitative techniques allow researchers to dig more deeply into the meanings of the data“ (Squire, 2005, S. 11)

(1) Das explorative Vorgehen ist zunächst darin begründet, dass über Usability-Anforderungen für die Zielgruppe der funktionalen Analphabeten und Analphabetinnen bisher in der wissenschaftlichen Literatur nicht berichtet wurde und

folglich solide Grundlagen für eine Hypothesenbildung und damit für die Beantwortung einiger Forschungsfragen (insbesondere der Forschungsfragen F1a und F1b) fehlen. Zudem sind neben der Erforschung der Usability-Anforderungen auch die Entwicklung und die Evaluation der Online-Testumgebung Bestandteile des Forschungsprojektes. Bestandteile der Untersuchung ist damit auch die Neuentwicklung eines Produktes, weshalb das explorative Vorgehen für die umfassende Evaluation unabdingbar ist, denn „Exploration spielt eine wichtige Rolle sowohl bei der Bildung wissenschaftlicher Theorien in der Grundlagenforschung als auch bei der Bildung technologischer Theorien in der angewandten Forschung (speziell: Evaluationsforschung)“ (Bortz & Döring, 2006, S. 355).

(2) Ein hypothesengeleitetes Vorgehen wird gewählt, weil Zusammenhänge überprüft werden sollen und die bestehende theoretische Basis die Formulierung von Hypothesen zulässt. Existierende Untersuchungen und Überlegungen erlauben die Vermutung über das Wirkungsgefüge zwischen der visuellen Gestaltung von Interfaces und der Beanspruchung des Arbeitsgedächtnis im Kontext computerbasierten Testens (vgl. Kap. 2.4 zur CLT und Kap. 2.6 zum Thema Usability).

Weitere methodische Vorüberlegungen begründen sich einerseits in den Forschungsfragen und -hypothesen als auch in den Erhebungsinstrumenten.

Methodische Vorüberlegung zu den Forschungsfragen und -hypothesen
Die Überprüfung der Hypothesen bedingten Vorüberlegungen hinsichtlich des Untersuchungsgegenstands, des Forschungsdesigns sowie der Stichprobengröße.
Untersuchungsgegenstand: Der Untersuchungsgegenstand ist die Online-Testumgebung otu.lea. Für die Überprüfung aller Hypothesen galt, dass jede Person sowohl die Dimension Lesen als auch die Dimension Schreiben bearbeitet. Die Dimensionen Lesen und Schreiben wurden gewählt, da die Aufgaben dieser Dimensionen im Vergleich zu den Dimensionen Mathematik und Sprachempfinden eine möglichst große Bandbreite an Aufgabenformaten (Single Choice, Textfelder, Korrekturformat) aufweisen und damit eine umfassende Evaluation hinsichtlich der Aufgabenformate gewährleistet ist. Diese Wahl ist zudem geeignet, um die Heuristiken „Interkulturelle Aspekte“ sowie „Joy of Use“ zu überprüfen, denn die gewählten Dimensionen weisen eine umfangreichere Einbettung des Storyboards vor (im Vergleich zu den Dimensionen Sprachempfinden und Mathematik). Zudem wurden verschiedene Aufgabenniveaus angeboten, um den heterogenen Kompetenzen der Teilnehmenden gerecht werden zu können und um die Probanden und Probandinnen weder zu über-

noch zu unterfordern⁸¹. Hierfür wurde innerhalb der Testumgebung zu Beginn des Tests von den Teilnehmenden anhand von Beispielen ihrer Kompetenzen eine Selbsteinschätzung für die Level-Auswahl gefordert. Die Bandbreite der Aufgaben ermöglichte in der Dimension Lesen eine Einteilung in drei Level (einfach, mittel und schwer), im Bereich Schreiben in zwei Level (einfach und mittel).

Forschungsdesign: Mit der Niveauewahl bestand zwar das Risiko, dass ein Großteil der Probanden und Probandinnen das gleiche Niveau wählt. Doch beeinträchtigt dies nicht die statistischen Berechnungen, da durch die Bildung relativer Performanzwerte eine Vergleichbarkeit aller Probanden und Probandinnen – unabhängig der Niveau- und Dimensionswahl – möglich ist: Die Performanz wurde mit Prozentwerten ausgegeben, so dass die Werte nicht die absolute Zahl der richtig bzw. falsch Antworten bezeichnen, sondern in Relation mit der Anzahl der bearbeiteten Aufgaben gesetzt wurden. Somit ist eine Vergleichbarkeit zwischen Dimensionen und Niveaus möglich (vgl. auch Kap. 8.1.2.1). Einer möglichen Verzerrung von Effekten durch eine inhaltliche Über- oder Unterforderung wurde durch die Selbsteinschätzung und der Niveauewahl entgegengewirkt und durch die Überprüfung auffälliger Werte und der Videoaufzeichnungen kontrolliert.

Stichprobengröße: Für die Überprüfung der Hypothesen zum *Extraneous Load* H1_{c1} („Es besteht ein Zusammenhang zwischen dem Einsatz von Cues und der Performanz: Der Einsatz von Cues hat einen positiven Einfluss auf die Performanz“) und H1_{c2} („Es besteht ein Zusammenhang zwischen dem Einsatz von Cues und der Bearbeitungszeit: Der Einsatz von Cues führt zu einer geringeren Bearbeitungszeit“) musste die Stichprobengröße auf mindestens 50 Personen festgelegt werden: Um die Effekte von Cues zu überprüfen, wurde ein Zwei-Gruppen-Design umgesetzt: Die Probanden und Probandinnen wurden randomisiert auf die Gruppen „Cues an“ und „Cues aus“ verteilt, so dass jeder Gruppe 25 Personen zugeordnet wurden.

Methodische Vorüberlegungen zu den Erhebungsinstrumenten

Die möglichen methodischen Probleme von Interviews sind Antwortverzerrungen, insbesondere die Zustimmungstendenz (Akquieszenz) und die soziale Erwünschtheit (Schnell & Hill, 2011, S. 346 ff). Die Zustimmungstendenz tritt insbesondere bei Personen mit geringer Ich-Stärke und bei unterprivilegierten Personen, die sich für den Alltag Behauptungsstrategien angeeignet haben, auf

⁸¹ Es war nicht möglich, die Teilnehmenden in Abhängigkeit ihres Kompetenzstands auszuwählen, da die Feststellung des Kompetenzstands einerseits für die Kursleitenden einen großen Aufwand bedeutet hätte sowie den Kursteilnehmenden gegenüber diskriminierend gewesen wäre und andererseits eine Mindestanzahl an Studienteilnehmenden nicht hätte gewährleistet werden können.

(Hare, 1960; Lenski & Leggett, 1960) (zitiert nach Schnell u. a., 2011, S. 347). Die soziale Erwünschtheit ist u. a. einerseits in dem Wunsch nach sozialer Anerkennung und andererseits in situationsspezifischen Reaktionen auf die Datenerhebung zu begründen. Zu vermuten war, dass funktionale Analphabeten und Analphabetinnen aufgrund ihrer literalen Defizite sowohl Behauptungsstrategien – insbesondere Vermeidungsstrategien – anwenden, als auch das Bedürfnis nach Anerkennung haben und daher in den Interviews sowohl die Gefahr der Akquieszenz als auch der sozialen Erwünschtheit besteht. Um dem Risiko entgegen zu wirken, wurde während der Interviewdurchführung um Erklärungen, Beispiele und Begründungen der Antworten gebeten. Grundsätzlich kann das Risiko der Antwortverzerrung allerdings nicht gänzlich kontrolliert werden, so dass dieses methodische Problem in der Ergebnisinterpretation und Methodenreflexion zu berücksichtigen ist.

Die Gefahr der Antwortverzerrungen wäre vermutlich höher, wenn tendenziell geschlossene Fragen zu den einzelnen Heuristiken gestellt werden würden. Daher waren das Forschungsdesign und der Interviewleitfaden dahingehend zu konzipieren, Spielraum für Aussagen über die bereits identifizierten Heuristiken hinaus zu ermöglichen und offene Fragen zu stellen. Damit sollten Aspekte aufgedeckt werden können, die nicht durch die bisherigen Erkenntnisse zum Thema Usability abgedeckt sind.

Eine weitere Herausforderung bestand in der Äquivalenz der Untersuchungsbedingungen. Während des Usability-Tests wurden den einzelnen Probanden und Probandinnen verschiedene Betreuer und Betreuerinnen zugewiesen. Je nach Probanden-/Probandinnenanzahl des jeweiligen Untersuchungssettings standen fünf bis acht Betreuer und Betreuerinnen zur Verfügung, so dass einer betreuenden Person eine Gruppe von bis zu drei Probanden und Probandinnen zugewiesen war. Die Betreuer und Betreuerinnen waren für die Unterstützung zuständig, wenn der Verlauf des Usability-Tests gefährdet war; beispielsweise wenn das Programm abstürzte oder eine Person aufgrund von Schwierigkeiten mit der Testumgebung drohte abbrechen. Die Unterstützung sollte sich lediglich auf Funktionen und nicht auf inhaltliche Aspekte der Kompetenzmessung richten, damit die Unterstützung nicht mit der Kompetenzdiagnostik interferiert und somit Datenqualität gewährleistet ist. Die Betreuenden wurden im Vorfeld entsprechend instruiert. Mit Hilfe der Videoaufzeichnungen wurden diese mögliche Abweichungen von den Instruktionen sowie Einflussfaktoren erfasst und kontrolliert.

8.2 Zyklus 3: Durchführung – Methodenwahl und Untersuchungskontext

Die Methodenwahl der Untersuchung begründet sich im Forschungsdesign, also in den Fragestellungen, den Hypothesen und dem Untersuchungskontext (wie

bereits im vorherigen Kap. 8.1.2.2 zu methodischen Vorüberlegungen dargestellt).

8.2.1 Forschungsdesign und Ablauf der Hauptuntersuchung

Die Hauptuntersuchung der Arbeit besteht aus einer umfassenden Usability-Studie über otu.lea. An der Studie nahmen sowohl funktionale Analphabeten und Analphabetinnen als auch eine Gruppe von Kursleitenden teil. Die funktionalen Analphabeten und Analphabetinnen absolvierten einen Usability-Test, der videographiert wurde und es wurden mit ihnen Befragungen und Interviews durchgeführt. Der Ablauf der Studie war für die funktionalen Analphabeten und Analphabetinnen institutionsübergreifend gleich. Mit der Gruppe von Kursleitenden wurde ein Experten-/Expertinneninterview durchgeführt. Das Interview mit Kursleitenden fand einmalig in einer Institution statt und wurde zeitlich parallel zu der Studie mit den funktionalen Analphabeten und Analphabetinnen durchgeführt. Die folgende Abbildung zeigt das forschungsstrategische Vorgehen der Usability-Studie des dritten Zyklus unter Berücksichtigung des zeitlichen Ablaufs.

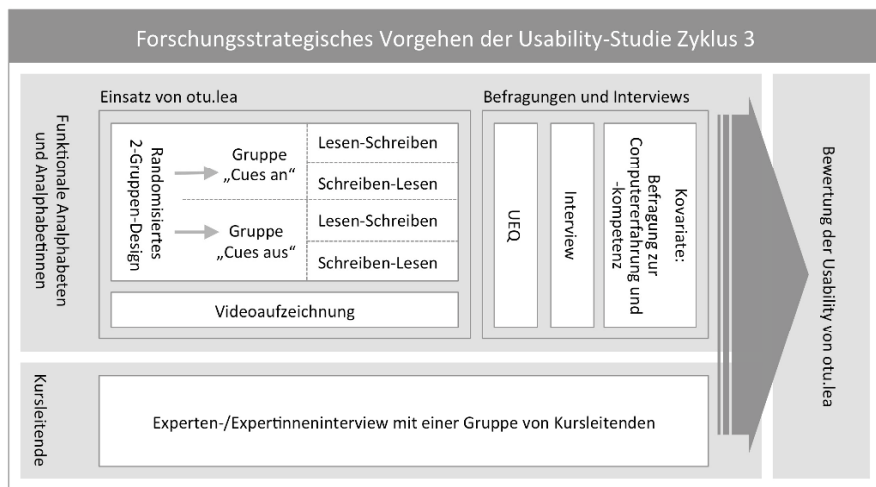


Abbildung 29: Forschungsstrategisches Vorgehen des dritten Evaluationszyklus

Die funktionalen Analphabeten und Analphabetinnen begannen mit dem Usability-Test. Für die Nutzung von otu.lea wurde die Gruppe der funktionalen Analphabeten und Analphabetinnen randomisiert auf zwei Gruppen aufgeteilt. Es wurden experimentelle Bedingungen eingefügt, woraus sich ein 2x2-

Versuchsplan ergibt. In Anlehnung an die vorherige Abbildung (Tabelle 20) stellt die folgende Abbildung den konkretisierten Versuchsplan hinsichtlich der experimentellen Bedingungen unter Angabe der voraussichtlichen Verteilung dar:

Reihenfolge	Lesen-Schreiben	Schreiben-Lesen
Cues		
Cues an	12 oder 13	12 oder 13
Cues aus	12 oder 13	12 oder 13

Tabelle 20: Experimentelle Bedingungen Usability-Studie 2013

Für die Hälfte der Nutzer und Nutzerinnen war das Blinken der Schaltflächen eingeschaltet („Cues ein“): Die Schaltflächen zur Ausführung von Aktionen – z. B. die Aktivierung des Lautsprecherbuttons für das Anhören von Instruktionen – waren in logischer Abfolge animiert, so dass die Animationen die idealtypische Reihenfolge für die Interaktionen zur Bearbeitung der Aufgaben vorgaben. Für die andere Hälfte waren die Animationen der Schaltflächen ausgeschaltet („Cues aus“). Die Verteilung auf die Gruppen erfolgte randomisiert hinsichtlich der experimentellen Bedingung sowie der Reihenfolge der Dimensionsbearbeitung (eine Gruppe bearbeitete zuerst die Dimension Lesen und anschließend die Dimension Schreiben, die andere Gruppe umgekehrt). Es war vorgesehen, dass jeder Proband/jede Probandin beide Dimensionen (Lesen und Schreiben) je einmal bearbeitet. Dem Proband/der Probandin wurde die Dimension, mit der er/sie beginnen sollte, vorgegeben. Das Niveau konnten sie selbst wählen (in der Abbildung 29 und Tabelle 20 nicht angegeben). Insgesamt bestand die Testcollection aus 191 Items. Von jeder Person konnten somit durchschnittlich ca. 2/5 bzw. ca. 78 der Items bearbeitet werden.⁸²

Anschließend beantworteten die Teilnehmenden den Fragebogen UEQ (Laugwitz u. a., 2006) zur UE von otu.lea. Darauf folgte die Durchführung des leitfadengestützten fokussierten Interviews. Abschließend wurden die Befragungen über die Computerkompetenz und -erfahrung der Probanden und Probandinnen durchgeführt.

Das Experten-/Expertinneninterview wurde mit einer Gruppe von Kursleitenden parallel zu der Erhebung mit den funktionalen Analphabeten und Analphabetinnen durchgeführt.

⁸² Die 191 Items verteilen sich auf fünf Niveaus – 3x Schreiben und 2x Lesen; von jeder Person sollte in jeder Dimension ein Niveau zur Bearbeitung gewählt werden – insgesamt sollte somit 2x ein Niveau pro Person gewählt werden. Daraus ergibt sich die Rechnung $191/5 \times 2$.

Vor dem Hintergrund der Ergebnisse aus den Erhebungen mit funktionalen Analphabeten und Analphabetinnen sowie den Kursleitenden wurde die Usability von otu.lea bewertet.

8.2.2 *Die Erhebungsinstrumente*

Die Forschungsfragen „Werden die als für funktionale Analphabeten und Analphabetinnen relevant identifizierten Heuristiken“ (F1a) und „Gibt es weitere Hinweise auf die Notwendigkeit, neue Heuristiken zu formulieren?“ (F1b) zum Thema Usability wurden über qualitative Interviews, anhand von Videoaufzeichnungen (Screen-Recording) des Usability-Tests sowie mit einem standardisierten Fragebogen zur UE exploriert. Die Fragen „Haben die Computererfahrung und –kompetenz einen Einfluss auf die Bedienbarkeit?“ (F2a) und „Haben die Computererfahrung und –kompetenz einen Einfluss auf die Performanz?“ (F2b) wurden mittels Data Logging (Bearbeitungszeit) sowie Fragebögen erhoben. Ob der Einsatz von Cues einen Effekt auf das Testergebnis und die Bearbeitungszeit (F3a: „Gibt es Hinweise auf die Beanspruchung des Extraneous Load durch den Einsatz von Cues?“) hat, wurde über das experimentell angelegte Design („Cues an“/„Cues aus“) untersucht. Weitere potenzielle Auslastungsfaktoren des Extraneous Load (F3b: „Gibt es Hinweise auf Faktoren, die den Cognitive Load im Allgemeinen und den Extraneous Load im Besonderen beeinflussen?“) wurden über Interviewaussagen ermittelt. Im Einzelnen werden die Instrumente im Folgenden vorgestellt.

8.2.2.1 Leitfadengestützte Interviews

Die Durchführung von Interviews zählt zu den benutzer-/benutzerinnenzentrierten Messansätzen der Usability-Evaluation (Niegemann, 2008). Es wurden fokussierte Interviews mit den jeweiligen Kursteilnehmenden sowie ein Gruppen-Interview mit Kursleitenden durchgeführt.

Das Vorgehen der Interviewplanung und der Erstellung eines Leitfadens richtete sich dabei nach Helfferich (2004). Für die Gewährleistung der intersubjektiven Nachvollziehbarkeit hat sie eine Liste von Forschungsentscheidungen erstellt⁸³. Deren wesentlichen Schritte werden nach einer kurzen Erläuterung auf die leitfadengestützten Interviews mit funktionalen Analphabeten und Analphabetinnen bezogen und sodann in einer kurzen Zusammenfassung auf das Experten-/Expertinnengespräch angewandt.

Erstellung des Leitfadens für das Interview mit funktionalen Analphabeten und Analphabetinnen

⁸³ Für eine umfassende Erläuterung s. Helfferich 2004 148 ff.

1) *Entscheidung für einen (präzisen) Forschungsgegenstand:* Nach dem Ansatz von Helfferich (2004) ist der Forschungsgegenstand sowohl inhaltlich als auch theoretisch-methodisch zu bestimmen, wobei die theoretisch-methodische Facette bestimmt, welchen Status die erhobenen Daten haben. Inhaltlich steht die Gebrauchstauglichkeit des Online-Diagnoseinstruments otu.lea im Fokus, theoretisch-methodisch geht es um die (problemzentrierte) Beschreibung eines gemeinsamen Erfahrungsgegenstandes.

2) *Entscheidung für eine Zielgruppe und Eingrenzung der Stichprobe:* Entsprechend der Forschungsfrage besteht die Zielgruppe aus funktionalen Analphabeten und Analphabetinnen, für welche die Usability des Diagnoseinstruments optimiert werden soll. Um zu erfassen, wie potenzielle Nutzer und Nutzerinnen die Usability bewerten, war die Evaluation entsprechend mit Angehörigen dieser Personengruppe durchzuführen (vgl. auch Kap. 2.7 zu Evaluationsmethoden Usability). Wie schon im Kap. 8.1.2.2 zu methodischen Voraussetzungen dargelegt, wurde die Stichprobe aufgrund der methodischen Bedingungen auf 50 Personen festgelegt. Diese im Kontext qualitativer Forschung vergleichsweise große Stichprobe ist darin begründet, dass eine sehr heterogene Nutzer-/Nutzerinnenschaft hinsichtlich der Ausdrucksfähigkeit, des Bildungsstands und der Computerkompetenz erwartet wurde: Die Teilnehmenden an Alphabetisierungskursen befinden sich auf unterschiedlichen Literalitätsniveaus und weisen ein Patchwork an Kenntnissen und Kompetenzen auf (Anslinger & Quante-Brandt, 2010).

3) *Entscheidung für eine Interviewform:* Die Entscheidung für eine Interviewform wird in Abhängigkeit der Forschungsfrage und der Zielgruppe bestimmt. Bestehen zu Beginn ein Vorverständnis des Untersuchungsgegenstandes seitens der Forschenden und die Definition relevanter Themenkomplexe, ist die Durchführung von Leitfadeninterviews angemessen. Die Ziele der Leitfaden-Interviews sind die Sammlung von Daten zu den Themenkomplexen sowie darüber hinaus zur Entwicklung gegenstandsbezogener Theorien (Friebertshäuser & Langer, 2010, S. 439). Letzteres ist wiederum auch ein explizites Ziel der DBR-Forschung (vgl. Edelson, 2002 und Kap. 3.1). Eine spezielle Form des Leitfaden-Interviews ist das fokussierte Interview. Das fokussierte Interview wurde ursprünglich für die Medienforschung entwickelt und hat zum Ziel, bestimmte Erfahrungen der Befragten thematisch konzentriert und detailliert zu erfassen (Friebertshäuser & Langer, 2010, S. 441). Hierfür ist ein gemeinsamer „Erfahrungsgegenstand“ Voraussetzung, zu dem Fragen bezüglich des subjektiven Empfindens, der Wahrnehmung und der Reaktionen gestellt werden. Zu beachten gilt hierbei, nicht-direktive Fragen zu stellen. Das Ziel ist nicht, die Generierung „objektiv“ interpretierbarer Ergebnisse im Sinne der klassischen Testtheo-

rie, sondern die subjektiven Erfahrungen zu erfassen und aus dem Interviewmaterial inhaltsanalytisch herauszuarbeiten.

„Die Inhaltsanalyse wird später nicht als die objektive Variante der Interpretation begriffen, sondern es interessieren die subjektiven Sichtweisen der dazu Befragten und ihr Verhältnis zum diskutierten Gegenstand [...]“ (Friebertshäuser & Langer, 2010, S. 441).

Zur Beantwortung der Forschungsfrage ist die spezielle Form des Leitfaden-Interviews, das fokussierte Interview, geeignet: Das – für die Durchführung von fokussierten Interviews notwendige – Vorverständnis über den Untersuchungsgegenstand generierte sich einerseits aus der Involviertheit in den Entwicklungsprozess und andererseits aus den durchgeführten Vorstudien in Form von Paper Prototyping-Experimenten (Zyklus 1), heuristischen Evaluationen und einer Usability-Studie (Zyklus 2). Hieraus konnten klare Problemfelder bzw. Themenkomplexe und ein klares Erkenntnisinteresse definiert werden, welches sich im Interviewleitfaden widerspiegelt. Mit der Usability-Studie sollten subjektive Erfahrungen der Probanden und Probandinnen mit otu.lea erfasst werden, nachdem sie otu.lea genutzt haben. Diese Erfahrungen bilden die Grundlagen für den gemeinsamen Erfahrungsgegenstand und erfüllen somit die Voraussetzungen für die Durchführung von fokussierten Interviews.

Ein weiteres Argument für die Verwendung von Leitfadeninterviews ist darüber hinaus die für qualitative Untersuchungen verhältnismäßig große Stichprobe. Der Leitfaden standardisiert zu einem gewissen Grad die inhaltliche Struktur und erleichtert zum einen die Auswertung; zum anderen ist dadurch eine bessere Vergleichbarkeit der Antworten möglich (Helfferich, 2004, S. 159).

4) *Erstellung eines Interviewleitfadens*: Der Interviewleitfaden ist auf der Grundlage der Theorie, des Untersuchungsgegenstandes und der Forschungsfrage zu erstellen, wofür drei Schritte vorgesehen sind (Helfferich, 2004, S. 163).

1. *Fragensammlung*: Auf der theoretischen Grundlage und in Abhängigkeit des Entwicklungsstands wurden Fragen gesammelt.
2. *Reduktion der Fragen unter Berücksichtigung des Vorwissens der Forschenden und der Offenheit*: Die Fragen wurden auf ihre Relevanz hin überprüft. Mit den beteiligten Programmierern und Programmiererinnen wurde erörtert, welche Aspekte des Diagnoseinstruments aus technischen Gründen klare Grenzen aufweisen, so dass eine Modifikation diesbezüglich (und somit auch die Exploration dieser Aspekte) nicht gewünscht ist und somit aus diesen Untersuchungen ausgeschlossen wurden – beispielsweise Fragen zur Datensicherheit. Des Weiteren

wurden die Fragen immer gemäß des aktuellen Entwicklungsstandes des Diagnoseinstruments gestellt.

3. *Überprüfung der Fragen bezüglich der methodischen Eignung:* Der im ersten Usability-Test (Zyklus 2 im Jahr 2012) eingesetzte Leitfaden wurde einerseits entsprechend der Erfahrungen und Ergebnisse des zweiten Zyklus modifiziert, so dass einige Fragen, die sich als zu stark geschlossen erwiesen, offener formuliert wurden. Andererseits wurden die Fragen inhaltlich an den Entwicklungsstand des Instruments angepasst. Beispielsweise war die Hilfefunktion zum Zeitpunkt des ersten Usability-Tests nicht vollständig implementiert, so dass diesbezüglich erst nach dem zweiten Usability-Test konkrete Fragen gestellt werden konnten. Die Fragen wurden mehrfach überprüft und im Vorfeld einem Pre-Test unterzogen.

Das oben genannte Vorgehen stellt damit sicher, dass die Fragen des Leitfadens einerseits auf der theoretischen Grundlage basieren, dem Entwicklungsstands angemessen sind sowie unter Berücksichtigung des Vorwissens entwickelt und hinsichtlich der methodischen Eignung überprüft wurden. Der Leitfaden für die fokussierten Interviews mit den Kursteilnehmenden umfasst folgende Fragen:

Einstiegsfragen:

1. Wie war es für Sie, das Programm zu nutzen?
2. Welche besonderen Erfahrungen haben Sie gemacht? Was hat Ihnen besonders Spaß gemacht?
3. Gab es Schwierigkeiten/Hatten Sie Probleme? Falls ja, konnten Sie diese lösen?

Erlernbarkeit:

4. Wie war der Anfang bzw. der Einstieg für Sie?
5. Welche Erfahrungen haben Sie mit der Hilfefunktion gemacht?

Kognitive Belastung:

6. Wie konnten Sie die gesprochenen Anweisungen verstehen (Ton/Inhalt)?
7. Was fällt Ihnen zum Aufbau/zur Struktur des Programms ein?
8. Erzählen Sie bitte, wie Sie sich zurechtgefunden haben. Gab es Hinweise, die Ihnen den Umgang mit dem Programm erleichtert haben?
9. Wünschen Sie sich mehr Informationen, wie das Programm zu nutzen ist? Wenn ja, an welchen Stellen?

Motivation:

10. Wie motiviert waren Sie, das Programm zu nutzen?
11. Was fällt Ihnen zu den Geschichten und den Figuren im Programm ein? Konnten Sie sich in die Personen hineinversetzen? Bevorzugen Sie für das Programm reale Fotos oder Zeichnungen?
12. Würden Sie gerne öfter mit diesem Programm arbeiten?
13. Fanden Sie die Aufgaben eher schwer oder eher leicht?

Design:

14. Wie gefällt Ihnen die farbliche Gestaltung?
15. Was ist Ihnen in Bezug auf die Schrift aufgefallen (ggf. Screenshot zeigen)?

Empfindung:

16. Wie fühlen Sie sich nach dem Test (z. B. erledigt, zufrieden, erschöpft, motiviert)?

Abschlussfrage:

17. Gibt es noch etwas, das Sie uns fragen möchten? Haben Sie Anmerkungen?

Abbildung 30: Interviewleitfaden fokussiertes Interview Kursteilnehmende Usability-Studie 2013

Entsprechend der Anforderungen an die Länge qualitativer Interviews (Helfferich, 2004, S. 159), ist die Anzahl der Fragen angemessen. Durch die Gliederung der Fragen in Themenbereiche ist der Leitfaden übersichtlich und orientiert sich an einem (idealtypischen) Erinnerungs-/Argumentationsfluss der Studienteilnehmenden, so dass die interviewten Personen den Fragen möglichst leicht folgen konnten. Zu Beginn wurde eine Einstiegsfrage formuliert, die möglichst offen gestellt ist und an dem Erfahrungshorizont der Probanden und Probandinnen anknüpfte.

Hinsichtlich der Durchführung der Interviews war unter Berücksichtigung der Zielgruppe zu beachten, die Fragen möglichst konkret und präzise zu stellen, um einer Überforderung durch einen zu großen Interpretationsspielraum seitens der interviewten Personen entgegen zu wirken⁸⁴. Darüber hinaus wurden Maßnahmen ergriffen, um einer Antwortverzerrung durch Suggestivfragen entgegen zu wirken. Die Befragungen in den ersten beiden Zyklen haben gezeigt, dass aufgrund von Verständnisproblemen manche Interviewer und Interviewerinnen dazu neigten, Leitfragen suggestiv zu formulieren und vertiefende Darstellungen von den interviewten Personen abgeblockt oder übergangen werden. Daher wurden im Vorfeld der Interviews eine umfangreiche Interviewschulung durch eine ausgewiesene Expertin in der qualitativen Forschung durchgeführt, an der alle Interviewer und Interviewerinnen teilgenommen haben. In dieser wurde die Interviewsituation geprobt und es wurden Übungen zur Vermeidung von Suggestivfragen durchgeführt⁸⁵.

Bei der Durchführung der Interviews wurden die üblichen ethischen Grundsätze nach Transparenz der Zielsetzung und Verwendung der Daten folgendermaßen gewährleistet: Nach der Erläuterung der Studie und der Absicht der Datenverwendung wurde eine Einverständniserklärung eingeholt. Die Daten wurden mit Hilfe von Personencodes anonymisiert, so dass das sog. Trennungs- oder Lösungsangebot besteht und das Datengeheimnis sichergestellt werden kann.

Erstellung des Leitfadens für das Interview mit Kursleitern und Kursleiterinnen
Für das Experten-/Expertinnengespräch ergibt sich in Anlehnung an die vier Schritte der Interviewplanung und -organisation nach Helfferich (2004) folgendes Vorgehen:

⁸⁴ Die Konkretisierung der Fragestellungen im Leitfaden stehen der Anforderung, bürokratisches Abhaken zu vermeiden, zum Teil entgegen. Dem mit präzisen Fragen einhergehenden Risiko des bürokratischen Abhakens (Helfferich, 2004) wurde durch spezifische Anweisungen der Interviewer und Interviewerinnen und entsprechenden Übungen in der Interviewer-/Interviewerinnenschulung (vgl. Kap. 8.2.2.1) begegnet.

1) *Entscheidung für einen (präzisen) Forschungsgegenstand:* Der Forschungsgegenstand ist die Prozessangemessenheit (das System sollte für die Erfüllung realer Aufgaben in typischen Einsatzfeldern optimiert sein) von otu.lea., d. h. das System sollte für die Erfüllung realer Aufgaben in typischen Einsatzfeldern optimiert sein. Es soll erforscht werden, in wie weit otu.lea aus der Perspektive der Kursleitenden in der Alphabetisierung eingesetzt und sowohl innerhalb als auch außerhalb von Kursen genutzt werden kann.

2) *Entscheidung für eine Interviewform:* Das Interview wurde in Form eines Experten-/Expertinnengesprächs (vgl. z. B. Bogner, Littig & Menz, 2005) mit Kursleitenden durchgeführt. Konkret handelt sich um ein leitfadengestütztes Gruppeninterview, in dem mehrere Experten und Expertinnen zur wahrgenommenen Prozessangemessenheit von otu.lea befragt und zu einem gegenseitigen Erfahrungsaustausch über die Alphabetisierungspraxis und Einsatzpotenziale computerbasierter Diagnostik angeregt wurden.

3) *Entscheidung für eine Zielgruppe und Eingrenzung der Stichprobe:* Das Gespräch fand mit fünf Kursleitenden aus unterschiedlichen Alphabetisierungseinrichtungen statt.

4) *Erstellung eines Interviewleitfadens:* Der Interviewleitfaden beinhaltet Fragen zur Alphabetisierungspraxis und zu potenziellen Nutzungsmöglichkeiten von otu.lea:

Einstiegsfragen:

1. Wie arbeiten Sie in den Alphabetisierungskursen?

Fragen zu Einsatz- und Nutzungsmöglichkeiten von otu.lea:

2. Haben Sie Erfahrungen mit dem Einsatz von Computern in Alphabetisierungskursen?
 - a. Falls ja: Welche Programme nutzen Sie?
 - b. Wie ist das für Sie, Computer einzusetzen? (Hilfreich, unterstützend, entlastend, stressig, aufregend...?)
3. Kennen Sie otu.lea?
 - a. Falls ja: Was fällt Ihnen zu otu.lea ein?
 - b. Würden Sie otu.lea einsetzen? Warum? Warum nicht?
 - c. Was brauchen Sie/die TN, damit Sie es gerne einsetzen?
4. Hätten Sie Lust, sich das Programm anzusehen, wenn es fertig entwickelt ist und uns eine Rückmeldung zu geben?

Abschlussfrage:

5. Gibt es noch etwas, das Sie uns fragen möchten? Haben Sie Anmerkungen?

Abbildung 31: Interviewleitfaden für das Experten-/Expertinnengespräch mit Kursleitenden

Verfolgte Zielsetzung bei dem Experten-/Expertinnengespräch war, dass sich durch die (Impuls-)Fragen ein Gespräch unter den Kursleitenden entwickelt, in dem Erfahrungen ausgetauscht und die Angemessenheit von otu.lea in der Alphabetisierungspraxis diskutiert werden, so dass die interviewende Person in dem Gespräch in den Hintergrund treten kann.

8.2.2.2 Videoaufzeichnung

Die Usability-Tests wurden videographiert, was der zusätzlichen Informationsgeneration hinsichtlich der nicht aufgabenbezogenen Bearbeitungszeit, des Betreuungsbedarfs und zur Überprüfung der Aussagekraft der quantitativen sowie qualitativen Daten dient. Denn eine mögliche Störvariable, die Einfluss auf die Performanz und das Nutzer- und Nutzerinnenverhalten während des Tests nehmen könnte, ist die Unterstützung durch die Betreuer und Betreuerinnen. Auch wenn die Betreuer und Betreuerinnen hinsichtlich der Unterstützung der Probanden und Probandinnen instruiert wurden, ab wann bzw. unter welchen Bedingungen sie in den Testverlauf eingreifen und welche Unterstützung sie geben können, konnte die Betreuungsleistung durchaus unterschiedlich und solcherart ausfallen, dass es das Ergebnis beeinflussen könnte. Mit Hilfe der Videos wird

dieser mögliche Einflussfaktor kontrolliert. Des Weiteren werden die Videos herangezogen, um Aussagen der interviewten Personen zu überprüfen und um weitere Informationen zum Testverlauf zu gewinnen. Aufgrund der Testlänge sowie der unterschiedlichen Konzentration- und Behaltensleistung war anzunehmen, dass sich die Probanden und Probandinnen an relevante Aspekte nicht erinnern. Die Videos sollen Erinnerungslücken kompensieren und weitere Hinweise auf Usability-Probleme liefern, welche über die Befragungen und Interviews nicht aufgedeckt werden können.

8.2.2.3 Befragungen

Es wurden Befragungen zur Computererfahrung und -kompetenz sowie zu den Erfahrungen mit otu.lea und dessen Wahrnehmung durchgeführt.

Computererfahrung und -kompetenz

Die Computererfahrung und -kompetenz wurden über Fragebögen erhoben. Sie dienten der Wissensgenerierung über den Befragungsgegenstand und der Einflusskontrolle der Computererfahrung/-kompetenz auf das Testergebnis.

Der Fragebogen über die Computererfahrung fragt nach Nutzungsgewohnheiten und insbesondere dem Nutzungsumfang und besteht primär aus geschlossenen Fragen (s. Kap. 1.2 im Anhang). Die Teilnehmenden schätzten auf einer fünfstufigen Likert-Skala ein, wie gerne sie mit dem Computer arbeiten (von „nicht gerne“ bis „sehr gerne“) und auf einer vierstufigen Likert-Skala (von „nie“ bis hin zu „mindestens 1x am Tag“) für welche Zwecke sie wie oft das Internet nutzen. Hinsichtlich des zuletzt genannten Themenspektrums umfasst die Skala lediglich vier Stufen, da davon ausgegangen wird, dass eine zu starke Differenzierung der Antwortmöglichkeiten eine unnötige Überforderung hervorrufen würde. Die hier gewählte vierstufige Antwortmöglichkeit ist ausreichend, um die Frage zu erfassen, ob die Probanden und Probandinnen überhaupt den Computer nutzen und wenn ja für was. Denn niedrige Computerkompetenzen von Probanden und Probandinnen haben, wie bereits dargestellt, das Potenzial die Datenlage zu verzerren. Aus Gründen der Effizienz wurde daher der Fokus auf das untere Ende der Skala hinsichtlich der Computerkompetenz und -erfahrung gelegt. In dem Fragebogen zur Computererfahrung sind des Weiteren demographische Hintergründe zu Geschlecht und Alter abgebildet.

Der Fragebogen zu Computerkompetenz (s. Kap. 1.1 im Anhang) wurde eingesetzt, um zu überprüfen, ob die computerbezogenen Voraussetzung für die Nutzung von otu.lea erfüllt sind. Hierbei wurden Fragen zur Eingabe einer Internetadresse in das entsprechende Textfeld, das Wissen über die Funktion der Maus sowie über die Hochstelltaste zur Großschreibung gestellt. Die Items beinhalten Bilder (von einer Website, einer Tastatur und einer Computermaus), an-

hand derer die Probanden und Probandinnen das ihrer Meinung nach richtige Verhalten auswählen konnten. Beispielsweise sollten sie anhand der Grafik einer Tastatur die Tastenkombination für die Großschreibung identifizieren.

User Experience Questionnaire

Der User Experience Questionnaire (UEQ) (Laugwitz u. a., 2006) ist ein Fragebogen zur Erfassung des subjektiv empfundenen Eindrucks, den Nutzer und Nutzerinnen in Bezug auf ein Produkt entwickelt haben (Rauschenberger, Thomaschewski & Schrepp, 2013, S. 72).

„[Der UEQ] zielt [...] auf eine schnelle, unmittelbare und möglichst unreflektierte Beurteilung eines Produktes ab“ (Laugwitz u. a., 2009, S. 220).

In dem UEQ werden semantische Differentiale (Gegensatzpaare wie beispielsweise „schnell – langsam“) verwendet, um die Qualität eines Testobjektes zu ermitteln. Beim UEQ handelt es sich um ein bereits empirisch validiertes Instrument, das in die faktorenanalytisch ermittelten Dimensionen Attraktivität, Effizienz, Durchschaubarkeit, Steuerbarkeit, Stimulation und Originalität eingeteilt ist. Insgesamt besteht der UEQ aus 26 bipolaren Items, die sich auf die sechs Dimensionen verteilen. Die Items sind in einer siebenstufigen Likert-Skala abgetragen (Rauschenberger u. a., 2013, S. 73). Für eine valide Datenbasis ist Voraussetzung, dass die semantischen Differentiale korrekt interpretiert werden müssen. Um den Fragebogen auch mit Personen bearbeiten zu können, sich nicht auf dem entsprechenden sprachlichen Niveau befinden, gibt es eine Version in leichter Sprache. Für die Risikovermeidung einer Fehlinterpretation und Überforderung der Probanden und Probandinnen, wurde eben dieser in einfacher Sprache eingesetzt. Der Einsatz bei Personen ohne ausreichende Sprachkenntnisse oder ausreichendes Sprachvermögen ist nicht geeignet. Idealerweise wird er direkt nach der Anwendung (z. B. im Rahmen eines Usability-Testes) – ohne dazwischenliegende Diskussionen, welche die eigentliche Meinung über das zu bewertende Produkt beeinflussen könnten – bearbeitet.⁸⁶

⁸⁶ Weitere Instrumente zur Erfassung der subjektiven Nutzer-/Nutzerinnenerfahrung sind z. B. der AttrakDiff und die System Usability-Scale (SUS). Die Entscheidung gegen den AttrakDiff ist darin begründet, dass dieser nicht in einfacher Sprache vorliegt und damit für die Zielgruppe ungeeignet ist. Der SUS wird ausgeschlossen, da der Fokus allein auf der Bewertung des Systems liegt und die Aussagekraft der Ergebnisse mit zehn Fragen relativ undifferenziert ausfällt. Der SUS gibt keinen Aufschluss darüber, warum die Teilnehmenden das System entsprechend bewerten. So wird der Einsatz dieses Instruments auch als Quick and Dirty-Methode bezeichnet.

8.2.2.4 Data-Logging mit otu.lea

Das Online-Diagnoseinstrument otu.lea ist nicht nur Evaluationsgegenstand, sondern zugleich auch Datenerhebungsinstrument. Mit otu.lea wurden die Lese- und Schreibkompetenzen erhoben. Den Teilnehmenden wurden Aufgaben verschiedener Schwierigkeitsstufen zur Auswahl gestellt, aus der sie einen Teil der Aufgaben bearbeiten sollten: Für die Dimension Lesekompetenz standen Aufgaben mit den Niveaus „einfach“, „mittel“ und „schwer“, für die Dimension Schreibkompetenz „einfach“ und „mittel“ zur Auswahl⁸⁷. Zudem wurde die Aufgabebearbeitungszeit geloggt, um neben den Kompetenzen zusätzliche Informationen über die Dauer der Aufgabebearbeitung und somit über mögliche Störquellen zu erhalten. Die Bearbeitungszeit wurde aufgabenbezogen (nicht itembezogen) ausgegeben. In der Speicherung der Bearbeitungszeit sind die Sequenz mit dem Einführungstutorial, die Login-Sequenz, die Dimensions- und Niveauewahl nicht berücksichtigt⁸⁸.

8.2.3 Zugang zur Stichprobe

Funktionale Analphabeten und Analphabetinnen sind außerhalb von Kursen und Institutionen nur sehr schwer als Testpersonen zu identifizieren, insbesondere wenn eine größere Gruppe anvisiert wird. Einen der wenigen Zugänge zu dieser Gruppe bilden Institutionen mit Alphabetisierungskursen. Für die Gewinnung der Stichprobe wurden sechs Institutionen angeschrieben. Vier Institutionen konnten für das Vorhaben gewonnen werden und nahmen schließlich an den Usability-Tests teil. Der Kontakt wurde über die Kursleitenden hergestellt. Diese stellten zudem die Raum- und Computerversorgung sicher. Für die Teilnahme wurden keine Bedingungen an literale Kompetenz und/oder Computererfahrung gestellt. Die Teilnehmenden erhielten als Anerkennung ihres Einsatzes eine Aufwandsentschädigung von 10 Euro pro Person.

8.2.4 Auswertungsmethoden

Die Auswertungsmethoden wurden entsprechend der jeweiligen Erhebungsmethode (qualitativ/quantitativ) gewählt. Die folgende Tabelle zeigt eine Übersicht über die Erhebungsinstrumente, den Datenoutput, die entsprechenden Auswertungsmethoden sowie den jeweiligen Themenbereich.

⁸⁷ Für die Dimension Schreiben standen nur zwei Niveaus zur Auswahl, da zum Zeitpunkt der Studie noch nicht alle Aufgaben funktionsfähig und implementiert waren. Die Anzahl der einsatzfähigen Aufgaben legitimierte eine Einteilung in zwei Niveaus.

⁸⁸ Die nichtaufgabenbezogene Bearbeitungszeit wird über die Videoaufzeichnungen ermittelt.

	Erhebungs- instrument	Daten- output	Auswertungs- methoden	Fokussierte(r) Themenbe- reich(e) und Heuristiken
Fokussiertes Interview	Leitfaden- interview	Qualita- tiv	Qualitative In- haltsanalyse	Heuristiken Aufgabenange- messenhaft, Erwartungskon- formität, Steuerbarkeit, Per- spektivübernahme, Pro- zessangemessenheit, Joy of Use, Interkulturelle Aspekte
Experten- /Expertin- engespräch/ Gruppeninter- view	Leitfaden- interview	Qualita- tiv	Qualitative In- haltsanalyse	Heuristik Prozessangemes- senheit
Videographie	Video	Qualita- tiv	Qualitative In- haltsanalyse	Heuristiken Erlernbarkeit, Selbstbeschreibungsfähigkeit
Performanz	otu.lea	Quanti- tativ	deskriptiv, infe- renzstatistisch	Performanz; Hinweise auf <i>Extraneous Load</i>
Data-Logging	otu.lea	Quantita- tiv	deskriptiv, infe- renzstatistisch	Hinweise auf <i>Extraneous Load</i> , Heuristik Wahrneh- mungssteuerung
UEQ	Fragebogen	Quantita- tiv	deskriptiv	Heuristiken Steuerbarkeit, Joy of Use
Computererfah- rung und -kompetenz	Fragebogen	Quantita- tiv	deskriptiv, infe- renzstatistisch	Hinweise auf <i>Extraneous Load</i>

Tabelle 21: Übersicht der Datenerhebungen des dritten Evaluationszyklus

Im Folgenden werden die Auswertungsmethoden anhand der Erhebungsmethoden (qualitativ und quantitativ) vorgestellt.

Auswertung der qualitativen Daten (Interviews und Videos)

Das Interview- und Videomaterial wird zur Beantwortung aller drei Fragebereiche (Usability-Heuristiken, Computererfahrung und –kompetenz sowie Beanspruchung des *Extraneous Load*) herangezogen. Die Interviewergebnisse sollten Aufschluss über die subjektive Bewertung und die Videos Zusatzinformationen über die Usability von otu.lea liefern. Für die Auswertung dieser Ergebnisse eignet sich die strukturierende qualitative Inhaltsanalyse nach Mayring (2010) sowie Mayring und Brunner (2010). Begründet ist dies darin, dass die Interviews einen eindeutig definierten Gesprächs- und die Videos einen gemeinsamen Beobachtungsgegenstand haben und das Forschungsvorhaben weder zum Ziel hat, das qualitative Material in seiner Gesamtheit umfassend und breit auszuwerten

(vgl. Globalauswertung nach Legewie, 1994), noch allein aus den Interviews und/oder den Videos Theorien zu überprüfen oder zu entwickeln (vgl. Grounded Theory Glaser, Strauss & Paul, 2010).

Entsprechend der qualitativen Inhaltsanalyse ist bei der Auswertung

- theoriegeleitet vorzugehen,
- ein Codierleitfaden bzw. ein Kategoriensystem zu entwickeln und in einer Überarbeitungsschleife zu modifizieren,
- nicht das Material isoliert, sondern auch der Kontext zu betrachten (Mayring & Brunner, 2010, S. 325 ff).

Aufgrund des theoretischen Vorwissens ist hierbei eine deduktive Kategorienbildung anzuwenden, indem von dem theoretischen Bezugssystem (s. Kap. 2) ausgehend Kategorien entwickelt werden. Bei diesem Verfahren werden theoriegeleitet Kategorien und ggf. Unterkategorien entwickelt, definiert und zueinander abgegrenzt (Mayring, 2000, S. 5). Ziel ist, eine eindeutige Zuordnung zum Material zu erreichen.⁸⁹ Ein schematischer Ablauf des Vorgehens sieht wie folgt aus:

⁸⁹ In Abgrenzung zu Kuckartz (2012) und Schreiber (2012) zeichnet sich die Auswertungsmethode nach Mayring insbesondere durch die Theoriegeleitetheit aus. Kuckartz und Schreiber betonen hingegen die Bedeutung von Kategorien am Material.

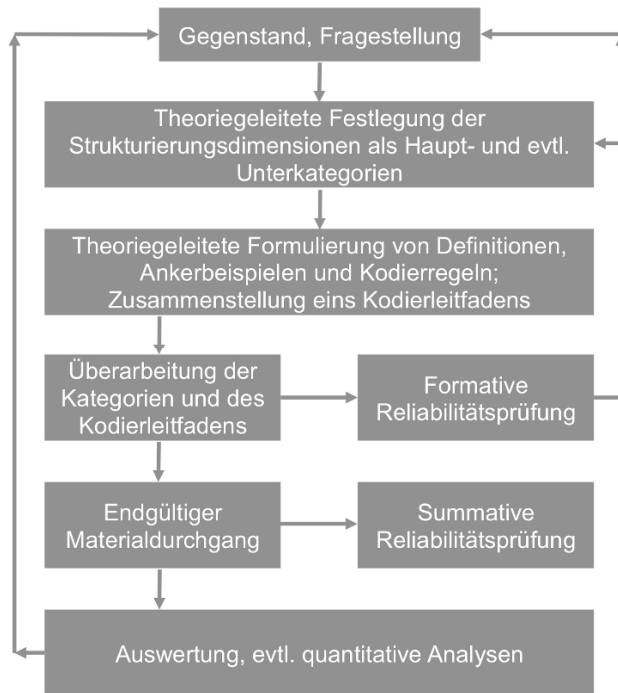


Abbildung 32: Ablauf deduktiver Kategorienbildung (Mayring, 2000, S. 5)

Gemäß des Ablaufmodells der deduktiven Kategorienanwendung wird zuerst die Fragestellung umrissen, die in einem zweiten Schritt mit den relevanten Theorien in Beziehung gesetzt und differenziert wird. Anschließend erfolgt in einem dritten Schritt die Definition von Ankerbeispielen, Codier- und Abgrenzungsregeln, woraufhin der Codierleitfaden zusammengestellt wird. In einem Zwischenschritt werden die Kategorien formal unter Berücksichtigung der Fragestellung überprüft (werden die interessierenden Aspekte unter Berücksichtigung der theoretischen Grundlage tatsächlich über die Kategorien abgebildet?). Fällt die Prüfung negativ aus, werden die Kategorien modifiziert. Sind diese Zwischenschritte für alle Kategorien abgeschlossen und steht das Kategoriensystem, erfolgt in einem fünften Schritt der Materialdurchgang. Dabei ist zudem eine summative Reliabilitätsprüfung, beispielsweise durch die Überprüfung der Intercoderreliabilität, durchzuführen. Schließlich wird im sechsten Schritt die qualitative (und ggf. quantitative) Auswertung vorgenommen. Formal ist dabei zu beachten, dass

sich jede Kategorie nur auf eine Bedeutungsdimension beziehen darf und sich die Kategorien einander ausschließen bzw. unabhängig voneinander sind (Schnell u. a., 2011, S. 400). Die Analyseschritte des Forschungsvorhabens werden entsprechend dieser Vorgehensweise umgesetzt und im Kapitel zur Analyse-Phase des dritten Zyklus` dargestellt.

Die Interpretation der Aussagen und Videobeobachtungen erfolgt dabei auf zwei Abstraktionsebenen: Auf der ersten Abstraktionsebene werden die Aussagen zusammengefasst und hinsichtlich ihres Aussagegehalts für die Online-Diagnostik interpretiert. Auf einer zweiten Abstraktionsebene werden die Aussagen mit den Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen in Beziehung gesetzt. Es wird zum einen überprüft, ob die Aussagen der funktionalen Analphabeten und Analphabetinnen sowie die Beobachtungen der Videoanalyse Hinweise darauf liefern, dass die Usability-Kriterien erfüllt werden. Zum anderen wird darüber hinaus exploriert, ob die Usability-Heuristiken den Bedürfnissen funktionaler Analphabeten und Analphabetinnen entsprechen. Ggf. können die Heuristiken wiederum auf der Grundlage der Usability-Studien modifiziert werden.

Auswertung der quantitativen Daten (Performanz, Computererfahrung und -kompetenz, UEQ)

Die Performanzdaten und die Ergebnisse zur Computererfahrung, zur Computerkompetenz sowie zur UE werden ebenfalls zur Beantwortung aller drei Fragebereiche herangezogen.

Die quantitativen Daten werden deskriptiv ausgewertet. Die Performanzdaten und Daten zur Computererfahrung/-kompetenz werden zusätzlich inferenzstatistisch analysiert, um die Hypothesen zu prüfen.

Die Ergebnisse des UEQ werden, für das Maß der internen Konsistenz, einer Reliabilitätsprüfung unterzogen. Der UEQ wird computergestützt mit Excel ausgewertet, da die Auswertungssyntax von den Entwicklern und Entwicklerinnen (Laugwitz u. a., 2006) des Fragebogens in Excel zur Verfügung gestellt wurde. Die weiteren quantitativen Daten werden computergestützt mit SPSS Version 22 ausgewertet.

8.2.5 Einhaltung der Gütekriterien

Die Einhaltung der Gütekriterien qualitativer und quantitativer Forschung werden im Folgenden kurz expliziert. Darüber hinaus sei auf die Kapitel zu methodischen Vorüberlegungen (Kap. 8.1.2.2) sowie zur Methodenwahl und Bestimmung des Untersuchungskontext (Kap. 8.2) verwiesen, in denen die Einhaltung der Gütekriterien am Beispiel des Forschungsvorhabens konkret zum Ausdruck kommt. Für das Forschungsvorhaben wird sich an den Gütekriterien qualitativer

Sozialforschung nach Steinke (2005) sowie den allgemeinen Gütekriterien der klassischen Testtheorie (Bortz & Döring, 2006) orientiert.

Die Kriterien für das Vorgehen qualitativer Forschung nach Steinke (2005:323 ff) werden im Folgenden aufgeführt und auf das Forschungsvorhaben bezogen:

Intersubjektive Nachvollziehbarkeit: Die intersubjektive Nachvollziehbarkeit wird durch die Dokumentation des Forschungsprozesses, die Überprüfung der InterCoder-Reliabilität sowie der Anwendung kodifizierter Verfahren durch die qualitative Inhaltsanalyse nach Mayring gewährleistet.

Indikation des Forschungsprozesses: Das qualitative Vorgehen und die Methodenwahl (Durchführung von fokussierten Interviews und Experten-/Expertinneninterviews sowie deren Auswertung mit der qualitativen Inhaltsanalyse) begründet sich in dem Ziel, die Usability von otu.lea zu erfassen und bewerten. Die Bewertung der Usability basiert dabei auf der Nutzungserfahrung der Probanden und Probandinnen mit otu.lea sowie auf deren subjektiver Sichtweise.

Empirische Verankerung: Für die Analyse der qualitativen Daten wird die qualitative Inhaltsanalyse nach Mayring angewandt, die es ermöglicht, über das Material einen Rückbezug zur Theorie herzustellen und darüber hinaus Offenheit für die Entwicklung neuer Theorien gewährt.

Limitation: Die Daten und deren Interpretation werden hinsichtlich ihres Geltungsbereichs und ihrer Generalisierbarkeit geprüft und diskutiert. Die qualitativen Gütekriterien spiegeln sich darüber hinaus auch in dem DBR-Ansatz wider, indem die intersubjektive Nachvollziehbarkeit, die Berücksichtigung aktueller Theorien und Forschungserkenntnisse (Indikation des Forschungsprozesses und empirische Verankerung) sowie die Überprüfung der Generalisierbarkeit Bestandteil des DBR-Prozesses sind (vgl. Kap.3).

Für das quantitative Vorgehen gelten die Gütekriterien Objektivität, Reliabilität sowie Validität (Bortz & Döring, 2006, S. 195 ff), die nachfolgend kurz expliziert und auf das Forschungsvorhaben bezogen werden.

Objektivität: Die Performanzdaten werden automatisiert erfasst und ausgewertet, so dass die Daten vom Testanwender unabhängig sind und damit eine hohe Objektivität gewährleistet ist. Um mögliche Einflüsse aufgrund von Unterstützungsleistung beim Lösungsprozess durch Betreuer und Betreuerinnen zu erfassen und zu kontrollieren, werden die Usability-Tests videographiert.

Reliabilität: Für die Erhebungen zur Computererfahrung/-kompetenz und der UE wird die Zuverlässigkeit der Messgenauigkeiten durch Bestimmungen des Zusammenhangsmaßes bzw. der internen Konsistenz anhand von Cronbachs α überprüft. Die Instrument zur Erhebung der Performanzdaten wurde bereits

empirisch validiert (Grotlüschen & Heinemann, 2011). Mögliche Einflussfaktoren aufgrund der Testmodalität werden über die Videos kontrolliert.

Validität: Die Inhaltsvalidität der Messung von Computererfahrung und –kompetenz wird über die theoretische Fundierung hergestellt (Kap. 2). Die Gültigkeit der Performanz- und der Bearbeitungszeitmessung wird durch die Berücksichtigung von möglichen Störvariablen bzw. Einflussfaktoren (z. B. die Computererfahrung) geprüft.

8.3 Zyklus 3: Analyse – Darstellung der Ergebnisse

Vorerst werden die allgemeinen deskriptiven Ergebnisse den Erhebungsinstrumenten zugeordnet dargestellt, da mitunter die gleichen Daten zur Beantwortung unterschiedlicher Fragen herangezogen werden. Beginnend mit demographischen Angaben, Angaben zur Bearbeitungszeit und Abbruchquote werden daraufhin die Daten der Befragungen zur Computererfahrung und -kompetenz, die Ergebnisse zur Performanz, zum UEQ und schließlich allgemeine Informationen zu den Interviews sowie zu den Videos dargestellt. Anschließend werden die Forschungsfragen mit Rückbezug auf die Ergebnisse bearbeitet und diskutiert.

8.3.1 Demographische Angaben, Bearbeitungszeit und Abbruchquote

An der Usability-Studie haben insgesamt 50 Personen teilgenommen. Die Personen verteilen sich über vier Institutionen in Norddeutschland. Die Probanden und Probandinnen waren zum Zeitpunkt der Erhebung Teilnehmende von Alphabetisierungskursen und 17 bis 57 Jahre alt. Das Durchschnittsalter betrug 36 (SD (Standardabweichung) = 13) und weist bei einer fast gleichmäßigen Verteilung über die Altersstufen eine große Streuung auf. 24 Personen sind männlich, 26 weiblich. Die Bearbeitungszeit mit otu.lea erstreckte sich von 12 bis hin zu 66 Minuten, der Mittelwert (MW) lag bei 31 Minuten (SD=13). Keine Person hat die Bearbeitung abgebrochen. Die gesamte Studie (Begrüßung, Durchführung des Usability-Test, der Interviews, Einsatz von Fragebögen) dauerte ca. 1,5 Stunden. Von 50 befragten Personen besitzen 36 Personen einen eigenen Computer.

8.3.2 Allgemeine deskriptive Ergebnisdarstellung

Den Erhebungsinstrumenten zugeordnet werden im Folgenden allgemeine deskriptive Ergebnisse dargestellt. Die Reihenfolge der allgemeinen Ergebnisdarstellung orientiert sich dabei an dem chronologischen Ablauf der Usability-Studie. Entsprechend werden zu Beginn die Ergebnisse zur Performanz (generiert durch das Data-Logging) und der Videographie dargestellt. Anschließend folgen die Ergebnisdarstellungen zur UE (erhoben mit dem UEQ), der Interviews sowie zur Computererfahrung und –kompetenz (erhoben mit Fragebögen).

Eine vertiefende Auswertung und Beantwortung der Fragestellungen findet ab Kap. 8.3.3 statt.

8.3.2.1 Allgemeine Ergebnisdarstellung zur Performanz

In dem Usability-Test wurden unterschiedliche Aufgabensets eingesetzt: Im Bereich Lesen standen Aufgabensets auf einfachem, mittlerem und schwierigem Niveau zur Auswahl. In der Dimension Schreiben konnte zwischen einem Aufgabenset auf dem einfachen und dem mittleren Niveau gewählt werden. Der Usability-Test des dritten Zyklus' zeigt folgende Verteilung:

Durchlauf	Cue	Dimen- sion	Niveau	n	Durchlauf	Cue	Dimen- sion	Niveau	n
1	an	L	leicht	4	2	an	L	leicht	4
1	an	L	mittel	6	2	an	L	mittel	3
1	an	L	schwer	2	2	an	L	schwer	0
1	aus	L	leicht	6	2	aus	L	leicht	0
1	aus	L	mittel	6	2	aus	L	mittel	5
1	aus	L	schwer	0	2	aus	L	schwer	2
1	an	S	leicht	7	2	an	S	leicht	3
1	an	S	mittel	6	2	an	S	mittel	5
1	aus	S	leicht	6	2	aus	S	leicht	6
1	aus	S	mittel	7	2	aus	S	mittel	4
Summe 50					Summe				32
					3	an	S	leicht	1
					Summe			1	

Tabelle 22: Experimentelles Design /Versuchsplan Ex-Post

Im ersten Durchlauf wurde 26-mal die Dimension Lesen und 24-mal die Dimension Schreiben bearbeitet. Jeweils bei der Hälfte der Personen waren die Cues eingeschaltet. In der Dimension Lesen wurde das einfache Niveau (1) zehnmal (4+6), das mittlere Niveau (2) zwölfmal (6+6) und das schwierige Niveau (3) zweimal gewählt. Im Bereich Schreiben wählten 13 Personen sowohl das einfache als auch das schwere Niveau (7+6 bzw. 6+7). Einen zweiten Durchlauf absolvierten 32 Personen, wobei 14 Personen die Dimension Lesen und 18 Personen die Dimension Schreiben wählten. Bei 15 Personen waren dabei die Cues aktiviert, bei 17 nicht. Vier Personen wählten im zweiten Durchlauf im Bereich Lesen das leichte, acht Personen das mittlere und zwei Personen das schwierige Niveau. Schreiben wählten beim zweiten Durchlauf jeweils neun Personen auf dem leichten und schwierigen Niveau an. Eine Person absolvierte einen dritten

Durchlauf im Bereich Schreiben auf dem leichten Niveau. Eine Person bearbeitete nur die Dimension Lesen (jeweils einmal einfach und einmal mittel). Eine Person absolvierte beide Durchläufe in der Dimension Schreiben. Dass nicht jede Person jede Dimension bzw. eine Dimension zweimal bearbeitet hat ist erstens in dem Zeitumfang für die Bearbeitung der ersten Dimension und zweitens in der Betreuung begründet.

Zu Erstens: Haben Personen mehr als 30 Minuten für eine Dimension benötigt, wurde kein weiterer Durchlauf begonnen. Insgesamt benötigten 15 Personen mehr als 30 Minuten für eine Dimension (zwei Personen bei Lesen-Einfach, eine Person im Bereich Lesen-Mittel, eine Person im Bereich Lesen-Schwer, vier Personen für Schreiben-Einfach und sieben Personen für Schreiben-Schwer). Organisatorisch und wegen möglicher Konzentrationsschwierigkeiten (aufgrund der langen Bearbeitungszeit) wurden die Probanden und Probandinnen direkt in das Interview gebeten, ohne eine zweite Dimension zu bearbeiten.

Zu Zweitens: Aufgrund dessen, dass nicht für jeden Untersuchungsteilnehmenden ein alleiniger Betreuer/eine alleinige Betreuerin zugeteilt war, ist es nicht gelungen, nach dem ersten Durchlauf bei jeder Person darauf zu achten, die Dimension zu wählen, die sie noch nicht bearbeitet hatte. Technisch hätte es zwar die Möglichkeit gegeben, dies zu kontrollieren. Da aber die Gefahr eines Programmabsturzes bestand wurde sich gegen eine technische Begrenzung entschieden⁹⁰.

Die Aufgaben bestanden aus insgesamt 180 Items⁹¹. Die Performanzdaten von Personen, bei denen das Programm zwischenzeitlich abstürzte, wurden zusammengeführt: Wenn eine Person bis zu einem Programmabsturz eine Aufgabensequenz von beispielsweise drei Aufgaben bearbeitet hatte, sie aufgrund des Programmabsturzes von vorne beginnen und die Sequenz von drei Aufgaben wiederholen musste, wurden die Performanzwerte des Durchlaufs vor dem Absturz in die Auswertung einbezogen. Gründe für die Einbeziehung der Performanzwerte vor dem Programmabsturz sind zum einen, den möglichen Lerneffekt auszuschließen und zum anderen, dass einige Probanden und Probandinnen bei dem zweiten Durchlauf die Aufgaben, die sie bereits im ersten Durchlauf vor

⁹⁰ Wäre das Programm zwischenzeitlich abgestürzt, hätte die Person mit der bisher noch nicht bearbeiteten Dimension fortfahren müssen, denn eine Dimension gilt aus der technischen Perspektive als bearbeitet, sobald lediglich eine Aufgabe bearbeitet wurde. Aus programmiertechnischen Gründen hätte im ungünstigsten Fall beispielsweise eine Person nur eine Aufgabe der Dimension Lesen bearbeitet und nach dem Programmabsturz mit der Dimension Schreiben fortfahren müssen. Um das zu vermeiden, fiel die Entscheidung für Kontrolle durch die Betreuer und Betreuerinnen.

⁹¹ Ein Item musste aufgrund einer defekten Audiodatei entfernt werden. Zwei Aufgaben im Bereich Schreiben konnten aufgrund eines fehlerhaften Auswertungsalgorithmus nicht in die Auswertung einbezogen werden, daher umfasst das Itemset nicht mehr wie in den methodischen Vorüberlegungen angemerkt 191 Items.

dem Absturz bearbeitet hatten, nicht beantworteten und die Aufgaben übersprangen. Für die Darstellung der Performanzwerte werden die Items aufgrund ihrer hohen Anzahl rechnerisch zusammengefasst und in Abhängigkeit der Kriterien Itemformat, Dimension und Niveau als Summenwerte sowie relative Performanzwerte dargestellt. In der Dimension Lesen wurden lediglich Single Choice-Items eingesetzt, so dass drei Kategorien zu bilden sind: Single Choice Lesen einfach, Single Choice Lesen mittel und Single Choice Lesen schwer. Im Bereich Schreiben wurden die Items in die Formatkategorien halboffen (z. B. Eingabe eines Buchstabens), offen (Eingabe eines Wortes oder Satzes in ein Textfeld) und Korrektur (Rechtschreibkorrektur eines Wortes) eingeteilt. So werden für Schreiben fünf Kategorien gebildet: Offenes⁹² Antwortformat Schreiben einfach, halboffenes Antwortformat Schreiben einfach, offenes Antwortformat Schreiben mittel, halboffenes Antwortformat Schreiben mittel sowie Korrekturformat Schreiben mittel. Die Korrekturitems wurden nur auf dem mittleren Niveau eingesetzt. Die Summenwerte sind in der folgenden Tabelle aufgeführt:

Format	Dimension/Niveau	N	Minimum	Maximum	MW	SD
Single Choice	Lesen einfach	14	4	20	14,79	4,26
	Lesen mittel	20	5	14	10,90	3,11
	Lesen schwer	24	5	15	8,50	4,73
Offen	Schreiben einfach	20	1	56	39,15	17,72
Halboffen	Schreiben einfach	23	2	25	17,22	6,77
	Schreiben mittel	22	3	16	9,09	3,75
Offen	Schreiben mittel	21	2	31	15,86	9,38
Korrektur	Schreiben mittel	21	0	3	1,95	1,24

Tabelle 23: Summenwerte format-, niveau- und dimensionsbezogen

Das Format „Single Choice Lesen einfach“ wurde 14-mal bearbeitet, „Single Choice Lesen mittel“ 20-mal und „Single Choice Lesen schwer“ 24-mal. Im Bereich Schreiben wurde 20-mal das Format „Offen Schreiben einfach“ und 23-mal „Halboffen schreiben einfach“ auf einfachem Niveau bearbeitet. Auf dem mittleren Niveau trat 22-mal das Aufgabenformat „Halboffen Schreiben mittel“, 21-mal „Offen Schreiben mittel“ und 21-mal das Format „Korrektur Schreiben mittel“ bearbeitet. In der Tabelle sind zudem Minimum, Maximum und Mittelwert der absoluten Performanzwerte (bzw. Summenwerte) dargestellt. Da die

⁹² Die Bewertung „offen“ ist in diesem Kontext relational zu den weiteren Itemformaten zu betrachten. Im theoretischen Diskurs über Itemformate werden offene Itemformate mitunter als Freitextaufgaben bezeichnet. Hier handelt es sich allerdings nicht um Freitextaufgaben, sondern um Textfelder, die im Vergleich zu den weiteren Items das am meisten geöffnete Format darstellen und nicht im Sinne des theoretischen Diskurses zu interpretieren sind.

Anzahl der Items der Kategorien unterschiedlich sind, sind diese für weitere Mittelwertvergleiche nicht verwendbar. Die Kategorien mit absoluten Performanzwerten können somit statistisch nicht in Relation zueinander gesetzt werden. Für Mittelwertvergleiche und für weitere Ergebnisdarstellungen werden daher für die Formatkategorien auch relative Performanzwerte in Prozent berechnet:

Format	Dimension/Niveau	N	Minimum	Maximum	MW	SD
Single Choice	Lesen einfach	14	20%	100%	74%	21%
	Lesen mittel	20	36%	100%	74%	20%
	Lesen schwer	24	00%	100%	65%	32%
Offen	Schreiben einfach	23	07%	76%	55%	19%
Halboffen	Schreiben einfach	20	02%	97%	70%	29%
	Schreiben mittel	22	25%	100%	60%	23%
Offen	Schreiben mittel	21	06%	91%	48%	26%
Korrektur	Schreiben mittel	21	00%	100%	65%	41%

Tabelle 24: Relative Performanzwerte format-, niveau- und dimensionsbezogen

Die Skalenkonsistenz wurde mit einer Reliabilitätsanalyse überprüft. Die Werte sind durchweg zufriedenstellend:

	Single Choice Lesen einfach	Single Choice Lesen mittel	Single Choice Lesen schwer	Halboffen Schreiben einfach	Offen Schreiben einfach	Halboffen Schreiben mittel	Offen Schreiben mittel	Korrektur Schreiben mittel
Ca	.85 (Items 19)	.78 (15 Items)	.88 (15 Items)	.89 (23 Items)	.98 (58 Items)	.82 (15 Items)	.94 (32 Items)	.82 (3 Items)

Tabelle 25: Skalenkonsistenz für die niveau- und dimensionsbezogenen Itemformate

Mit einem Cronbachs α ⁹³ zwischen .78 und .98 weist die Skalenkonsistenz zufriedenstellende Werte auf und erfüllt somit das Kriterium für weitere Berechnungen mit den Werten der Formatkategorien⁹⁴.

8.3.2.2 Allgemeine Ergebnisse der Videoanalyse

Die Videos werden nach der qualitativen Inhaltsanalyse nach Mayring ausgewertet. Folgende Kategorien wurden gebildet:

⁹³ Da es sich bei den Items um dichotome Items handelt, wurde zur Berechnung der Skalenwerte die Kuder-Richardson-Formel verwendet.

⁹⁴ Weitere Voraussetzungen für die Berechnungen, wie z. B. Normalverteilung, sind ebenfalls gegeben.

Nr.	Kategorie	Erläuterung
1.	Unterstützung	
1.1	unaufgeforderte Unterstützung funktional	Der Betreuer/Die Betreuerin unterstützt den Proband/die Probandin, ohne dass er/sie die Unterstützung eingefordert hat hinsichtlich Funktionalitäten.
1.2	unaufgeforderte Unterstützung inhaltlich	Der Betreuer/Die Betreuerin unterstützt den Probanden/die Probandin, ohne dass er/sie die Unterstützung eingefordert hat hinsichtlich der inhaltlichen Bearbeitung. Die inhaltliche Bearbeitung bezieht sich auf Unterstützungsleistung, die zur Lösung der Aufgaben führt und/oder der Erläuterung von Funktionen, die sich direkt auf die Performanz auswirkt.
1.3	eingeforderte Unterstützung funktional	Der Proband/Die Probandin fordert Unterstützung hinsichtlich der Funktionalitäten, die sich nicht direkt auf die Performanz auswirken.
1.4	eingeforderte Unterstützung inhaltlich	Der Proband/Die Probandin fordert Unterstützung hinsichtlich der inhaltlichen Bearbeitung. Die inhaltliche Bearbeitung bezieht sich auf Unterstützungsleistung, die zur Lösung der Aufgaben führt und/oder der Erläuterung von Funktionen, die sich direkt auf die Performanz auswirkt.
1.5	Bestätigungsunterstützung	Der Proband/Die Probandin fordert von dem Betreuer/der Betreuerin die Bestätigung der Richtigkeit seiner/ihrer Aktionen.
2.	Nutzung der Hilfefunktion	Der Proband/Die Probandin nutzt die Hilfefunktion.
3.	Login	
3.1	Login mit Unterstützung	Der Proband/Die Probandin benötigt Unterstützung beim Anmelden.
3.2	Login ohne Unterstützung	Der Proband/Die Probandin benötigt keine Unterstützung beim Anmelden.
3.3	kein Login beobachtbar	Es konnte nicht beobachtet werden, ob der Proband/die Probandin beim Anmeldung Unterstützung bekam.
4.	technische Schwierigkeiten	Es treten technische Schwierigkeiten auf.
5.	besondere Beobachtung	Beobachtungen, die nicht durch die übrigen Kategorien abgedeckt sind aber für die Beantwortung der Fragestellung relevant sind.

Tabelle 26: Auswertungsleitfaden Videoanalyse

Die Videos dienen primär der Beurteilung, ob otu.lea selbstständig nutzbar ist. Daher liegt der Fokus auf Unterstützungsfunktionen (Unterstützung durch Betreuer und Betreuerinnen sowie durch Nutzung der Hilfefunktion). Technische Schwierigkeiten werden codiert, um Ursachen für mögliche Probleme zu identi-

fizieren bzw. um Ergebnisse – beispielsweise zur Bearbeitungszeit – validieren und um diese Kategorie als Einflussfaktor auf die weiteren Ergebnisse (Performanz, Bearbeitungszeit) kontrollieren zu können. Zudem werden besondere Beobachtungen codiert, die keine eindeutige Zuordnung zu den gebildeten Kategorien ermöglichen aber Hinweise auf die Usability von otu.lea liefern.

Die rein quantitativ-deskriptive Auszählung der Videoanalyse ist in der folgenden Tabelle dargestellt.

	kein mal	weniger als 5 mal	weniger als 10 mal	mehr als 10 mal	Gesamt
unaufgeforderte Unterstützung inhaltlich	40	8	2	0	50
unaufgeforderte Unterstützung funktional	23	17	5	5	50
eingeforderte Unterstützung inhaltlich	33	14	1	2	50
eingeforderte Unterstützung funktional	22	22	3	3	50
Nutzung der Hilfefunktion	42	8	0	0	50
Technische Schwierigkeiten	33	17	0	0	50

Tabelle 27: Quantitative Auswertung der Videobeobachtungen

40 Probanden und Probandinnen wurden hinsichtlich der inhaltlichen Bearbeitung nicht unterstützt, acht Personen weniger als fünfmal und zwei Personen weniger als zehnmal. Hinsichtlich der Funktionalität erhielten 23 Personen keine Unterstützung, während 17 Personen weniger als fünfmal, fünf Personen weniger als zehnmal und fünf Personen mehr als zehnmal unaufgefordert Unterstützung zu Funktionalitäten erhielten. Eingefordert wurde die inhaltliche Unterstützung von 33 Probanden und Probandinnen gar nicht, von 14 Personen weniger als fünfmal, weniger als zehnmal von einer Person und mehr als zehnmal von zwei Personen. Funktionale Unterstützung wurde von 28 Personen nicht eingefordert. 22 Personen forderten Unterstützung hinsichtlich der Funktionalitäten ein, drei Personen taten dies weniger als zehn- und drei Personen mehr als zehnmal. Die Hilfefunktion wurde von acht Personen genutzt, wobei vier Personen die Hilfefunktion einmal, drei Personen dreimal und eine Person viermal aufrief. Technische Schwierigkeiten traten bei 17 Personen jeweils weniger als fünfmal auf. Technische Schwierigkeiten äußerten sich im Absturz des Programms und in einer nicht funktionierenden Audiodatei innerhalb einer Textaufgabe. Bei 30 Personen wurden besondere Beobachtungen, die über die definierten Kategorien hinausgehen, gemacht. Diese werden in der fragenbezogenen Ergebnisdarstellung vorgestellt und erläutert.

8.3.2.3 Allgemeine Ergebnisdarstellung zum User Experience Questionnaire

An der Befragung mit dem UEQ haben 50 Personen teilgenommen. Von den 26 Items wurde das Item mit dem Gegensatzpaar „effizient-ineffizient“ im Vorfeld entfernt, da davon ausgegangen wurde, dass das Item eine zu hohe Schwierigkeit aufweist. Von den sechs „Skalen“ (Attraktivität, Durchschaubarkeit, Effizienz, Steuerbarkeit, Stimulation und Originalität (Laugwitz u. a., 2006) weisen nur zwei ein mäßig zufriedenstellendes Cronbachs α auf: Attraktivität (Cronbachs $\alpha=.77$) und Stimulation (Cronbachs $\alpha=.69$). Da es sich beim UEQ um ein validiertes Instrument handelt ist davon auszugehen, dass die niedrigen Skalenkonsistenzen der Skalen Durchschaubarkeit, Effizienz, Steuerbarkeit und Originalität vermutlich in einer fehlerhaften Interpretation der Gegensatzpaare dieser Skalen begründet sind. Somit sind die Antworten zu diesen Items nicht aussagekräftig und können in der weiteren Ergebnisdarstellung und Interpretation nicht berücksichtigt werden (die Ergebnisse im Einzelnen sind im Anhang in den Tabellen 4-7 dargestellt). Die Angaben zur Attraktivität sind in folgender Tabelle anhand der positiv konnotierten Ausprägungen der Gegensatzpaare dargestellt:

		erfreulich	gut	anziehend	angenehm	schön	sympathisch
N	Gültig	50	50	46	49	50	49
	Fehlend	0	0	4	1	0	1
Mittelwert		1,76	2,58	1,91	2,63	2,36	2,35
Standardabweichung		1,76	1,03	1,40	,73	1,41	1,22
Varianz		3,08	1,07	1,95	,53	1,99	1,48
Minimum		-3	-2	-2	0	-3	-2
Maximum		3	3	3	3	3	3
* Cronbachs $\alpha =.77$							

Tabelle 28: Ergebnisse des UEQ für die Skala Attraktivität

Die Skala beinhaltet die Werte von -3 bis +3. Das Attribut „erfreulich“ wurde durchschnittlich mit 1,76 (SD=1,76) und „gut“ im Durchschnitt mit 2,58 (SD=1,03) beurteilt. Die Probanden und Probandinnen bewerteten otu.lea durchschnittlich mit 1,91 (SD=1,4) als anziehend und mit 2,63 (SD=0,73) als angenehm. Zudem nahmen die Probanden und Probandinnen otu.lea mit einem Mittelwert von 2,36 (SD=1,41) als schön und mit einem Mittelwert von 2,35 (SD=1,22) als sympathisch wahr. Insgesamt wird die Attraktivität von otu.lea mit einem Mittelwert von 2,27 als ausgeprägt wahrgenommen.

Die Skala Stimulation besteht aus vier Attributen, die in der nachstehenden Tabelle aufgeführt sind:

		erfrischend	spannend	inter- sant	abwechslungs- reich
N	Gültig	50	50	49	47
	Fehlend	0	0	1	3
Mittelwert		1,80	1,78	2,43	1,60
Standardabweichung		1,37	1,58	1,02	1,81
Varianz		1,88	2,50	1,04	3,29
Minimum		-1	-3	-1	-3
Maximum		3	3	3	3
* Cronbachs α =.69					

Tabelle 29: Ergebnisse des UEQ für die Skala Stimulation

Die Stimulation wird anhand der Attribute erfrischend (Mittelwert 1,8, SD=1,37), spannend (Mittelwert 1,78, SD=1,58), interessant (Mittelwert 2,43, SD=1,02) sowie abwechslungsreich (Mittelwert 1,6, SD=1,81) bewertet. Der Mittelwert der einzelnen Attribute und somit der Wert für die Skala Stimulation beträgt 1,9 und kann ebenfalls als ausgeprägt beurteilt werden.

8.3.2.4 Allgemeine Ergebnisse der Interviews

Analysebestandteil sind die vorliegenden Interviews von Teilnehmenden der Usability-Studie und das Experten-/Expertinnengespräch. Zwei Interviews mit funktionalen Analphabeten/Analphabetinnen wurden aufgrund technischer Probleme nicht mit einem Aufnahmegerät aufgezeichnet und werden daher in der Analyse nicht berücksichtigt.

Der strukturierenden Inhaltsanalyse entsprechend wurden deduktiv Kategorien gebildet, mit Abgrenzungsregeln sowie Ankerbeispielen versehen und schrittweise am Material modifiziert. Der daraus generierte Auswertungsleitfaden für die fokussierten Interviews mit Kursteilnehmenden beinhaltet folgende Kategorien:

Nr.	Kategorie	Erläuterung
1.	Erlernbarkeit	
1.1	Bewertung Einstieg	Empfindung/Wahrnehmung des Einstiegs (Video & Eingeben des Anfangs-Codes); Benennung von Schwierigkeiten
1.2	Nutzung Hilfefunktion	Nutzung der Hilfefunktion (Fragezeichen) allgemein; Beschreibung an welcher Stelle sie genutzt wurde und Begründung für die Nutzung; Nichtnutzung der Hilfefunktion
1.3	Nutzung anderer Hilfefunktionen	Nutzung von Funktionen, die als Hilfe empfunden wurden nicht aber der Hilfefunktion zuzuordnen sind (z. B. Lautsprecher,
1.4	Schwierigkeiten/Probleme	Erläuterung von Schwierigkeiten und Problemen während der Nutzung von otu.lea, ausgenommen sind
2.	Kognitive Belastung	
2.1	Verständnis Ton	Kommentare zur Verständlichkeit des Tons (Lautstärke/Stimme/Sprechtempo)
2.2	Verständnis Inhalt	Verständnis des Inhalts (Aufgabeninstruktion/Aufgabeninhalt)
2.3	Aufbau/Struktur	Kommentare zur Struktur/dem Aufbau von otu.lea
2.4	Orientierung	Erläuterungen, wie sich zurechtgefunden wurde
2.5	Hilfreiche Informationen	Erwähnung, ob es Informationen gab, die zum weiteren Nutzungsverlauf geholfen haben
3.	Motivation	
3.1	Motivation der Nutzung (Anfangs und am Ende)	Beschreibung der selbst wahrgenommenen Motivation, das Programm anfangs zu nutzen und ob sich für eine Weiterempfehlung bzw. eine weitere Nutzung ausgesprochen wurde. Einbezogen werden auch Aussagen zur generellen Zufrieden-
3.2	lea.-Charaktere	Kommentare zu den lea.-Charakteren und dem Storyboard, die nicht die Identifikation oder die Bewertung der Zeichnun-
3.3	Identifikation mit lea.-Charakteren	Aussagen zum Identifikationspotenzial mit den lea.-Charakteren
4.	Design	
4.1	Bewertung des Designs: Farbe	Kommentare zur farblichen Gestaltung
4.2	Bewertung des Designs: Schrift	Kommentare zur Schriftart
5.	Empfindung	Äußerungen zur selbst wahrgenommenen Befindlichkeit
6.	Verbesserungsvorschläge	Aussagen zu konkreten Verbesserungsvorschläge

Tabelle 30: Auswertungsleitfaden für die fokussierten Interviews mit Kursteilnehmenden

Die Interviews dauerten durchschnittlich 00:09:34 Minuten (SD=00:04:08), wobei das längste Interview 19:49 Minuten und das kürzeste Interview 3:44 Minuten geführt wurde. Um die InterCoder- (oder auch Interrater-) Reliabilität zu messen, wurden drei Interviews von zwei unabhängigen Ratern codiert. Die InterCoderreliabilität wurde a) anhand der prozentualen Übereinstimmung für die Häufigkeiten der Codierungen mit der Datenanalysesoftware MAXQDA, b) anhand der prozentualen Übereinstimmung der codierten Segmente nach Cohens Kappa sowie c) anhand des Übereinstimmungsmaßes Cohens Kappa κ unter Berücksichtigung der Kategorienanzahl berechnet. In folgender Tabelle sind die Ergebnisse der Reliabilitätsprüfungen zusammengefasst:

Interviews	D3	V3	AV3
Reliabilitätsprüfung			
a) Prozentuale Übereinstimmung	56%	60%	60%
b) Prozentuale Übereinstimmung nach Cohens Kappa	33%	75%	63%
c) Cohens Kappa κ	.29	.73	.60
Korrigierter Cohens Kappa κ	.79	/	/

Tabelle 31: Ergebnisse der Reliabilitätsprüfung

Für das Interview D3 liegt sie bei 56%, für das Interview V3 und AV3 bei 60%⁹⁵. Zusätzlich wurde der Cohens Kappa berechnet⁹⁶. Für die Berechnung des Cohens Kappa ist die prozentuale Übereinstimmung der codierten Segmente (d. h. es wird Anfang und Ende des codierten Abschnitts in die Bewertung einbezogen und nicht nur, ob ein Code vergeben wurde) ein zentraler Wert, der ebenfalls mit Hilfe von MAXQDA berechnet wurde. Die Fehlertoleranz wurde auf 10% festgelegt. Demnach dürfen die codierten Segmente um 10% voneinander abweichen. Dies ist ein recht strenger Wert. Die prozentuale Übereinstimmung beträgt 33% (D3), 75% (V3) sowie 63% (AV3). Der Cohens Kappa nach Brennan und Prediger (1981) beinhaltet ein zufallskorrigiertes Maß, indem die zufällige Übereinstimmung anhand der Kategorienanzahl berücksichtigt wird ($1/\text{Anzahl der Kategorien}$). Für das Interview D3 liegt der Cohens Kappa bei $\kappa=.29$, für das Interview V3 bei $\kappa=.73$ und für das Interview AV3 bei $\kappa=.60$. Bei einer näheren Überprüfung der Codierungen des Interviews D3 sind zwei Gründe für den niedrigen Wert zu identifizieren: 1) Im letzten Drittel des Interviews sind die Zeilennummern nicht mehr identisch, da ein Absatz eingefügt wurde. Da die Zeilennummern aber für die Berechnung der Übereinstimmung ausschlaggebend sind, hat eine nicht identische Zeilennummerierung einen großen

⁹⁵ Kriterium für diese Berechnung ist die Häufigkeit der Zuordnung des Codes im Dokument.

⁹⁶ Der Cohens Kappa berechnet sich aus $\text{Kappa} = [\text{Gemessene Übereinstimmung} - 1/k] / [1 - 1/k]$. „k“ bezeichnet die Anzahl der Kategorien (in diesem Fall 18).

Effekt auf die Berechnung der InterCoderreliabilität. 2) Die Codierweisen beider Codierer/Codiererinnen sind unterschiedlich: Einerseits hat ein Codierer/eine Codiererin inhaltlich zusammenhängende Textpassagen mit einzelnen Codierungen versehen wohingegen der andere Codierer/die andere Codiererin inhaltlich zusammenhängende Passagen „am Stück“ mit einem Code versehen hat. Andererseits wurden von einem Codierer/einer Codiererin Wiederholungen codiert, wohingegen der zweite Codierer/die zweite Codiererin das nicht tat. Wird das Interview entsprechend überarbeitet (Vereinheitlichung der Zeilen, Zusammenfassung von Codierungen einzelner inhaltlich zusammenhängender Textstellen mit dem gleichen Code und Vereinheitlichung hinsichtlich der Wiederholungen) ergibt der daraufhin korrigierte Cohens Kappa für das Interview D3 den Wert von $\kappa=.79$. Wird die Bewertung nach Fleiss und Cohen (1973) zugrunde gelegt, ist der Cohens Kappa κ für die zwei Interviews D3 und V3 als mindestens gut, für das Interview AV3 als zufriedenstellend zu bewerten⁹⁷.

Das Experten-/Expertinnengespräch mit den Kursleitenden fand an einer Volkshochschule statt, an der zeitgleich eine Usability-Studie durchgeführt wurde. Teilgenommen haben fünf Kursleitende, davon vier Frauen und ein Mann. Das Gespräch dauerte 34 Minuten und 12 Sekunden. Für die Auswertung nach der strukturierenden Inhaltsanalyse wurden folgende Kategorien gebildet:

Nr.	Kategorie	Erläuterung
1	Potenzielle Nutzung otu.lea anonym	Aussagen zur möglichen Nutzung von otu.lea außerhalb der Alphabetisierungskurse
2	Potenzielle Nutzung otu.lea Alpha-Kurse	Aussagen zur möglichen Nutzung von otu.lea in Alphabetisierungskursen
3	Passung otu.lea und Alpha-Kurse	Hinweise auf eine Passung zwischen (der Gestaltung von) otu.lea und den speziellen Anforderungen in der Alphabetisierungsarbeit
4	Sonstige Bemerkungen	Weitere Bemerkung, die Hinweise auf die zielgruppen- und kursangemessene Nutzung von otu.lea liefern

Tabelle 32: Auswertungsleitfaden für das Experten-/Expertinnengespräch mit Alphabetisierungskräften

Das Experten-/Expertinnengespräch wird ausschließlich zur Beurteilung der Heuristik Prozessangemessenheit herangezogen⁹⁸.

⁹⁷ Nach Fleiss und Cohen (1973) ist ein Cohens Kappa zwischen $\kappa=.60$ und $\kappa=.70$ als gut zu bewerten.

⁹⁸ Aufgrund der wenigen Kategorien wurde auf eine Überprüfung der InterCoder-Reliabilität und der Formulierung von Abgrenzungsregeln verzichtet.

8.3.2.5 Allgemeine Ergebnisse zur Computererfahrung und -kompetenz

Die Computererfahrung und -kompetenz wurde mit Fragebögen im Anschluss an den Usability-Test erhoben. Die Fragen wurden den Teilnehmenden vorgelesen (s. Fragebögen „Computerkompetenz“ und „Computererfahrung“ im Anhang).

Computererfahrung

Die Computererfahrung wurde über die Einstellung (Wie gerne arbeiten Sie mit dem Computer?), den Computerbesitz, die Nutzungsdauer sowie über die Internetnutzung erhoben.

Wie gerne arbeiten Sie mit dem Computer?

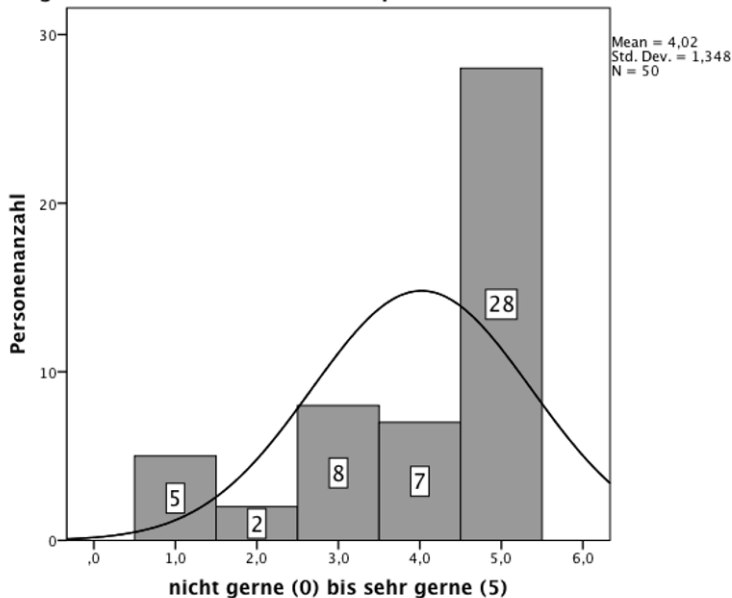


Abbildung 33: Positive Einstellung zur Computerarbeit

28 Personen (56%) arbeiten sehr gerne und sieben Personen (14%) gerne mit dem Computer. Mit „weder noch“ antworteten acht Personen (16%). Zwei Personen (4%) arbeiten tendenziell ungern und fünf Personen (10%) nicht gerne mit dem Computer. Mit 35 Personen (70%) ist somit ein Großteil gegenüber der Nutzung von Computern positiv eingestellt. Die Nutzungsdauer zeigt hingegen, dass viele Personen tendenziell wenig den Computer nutzen:

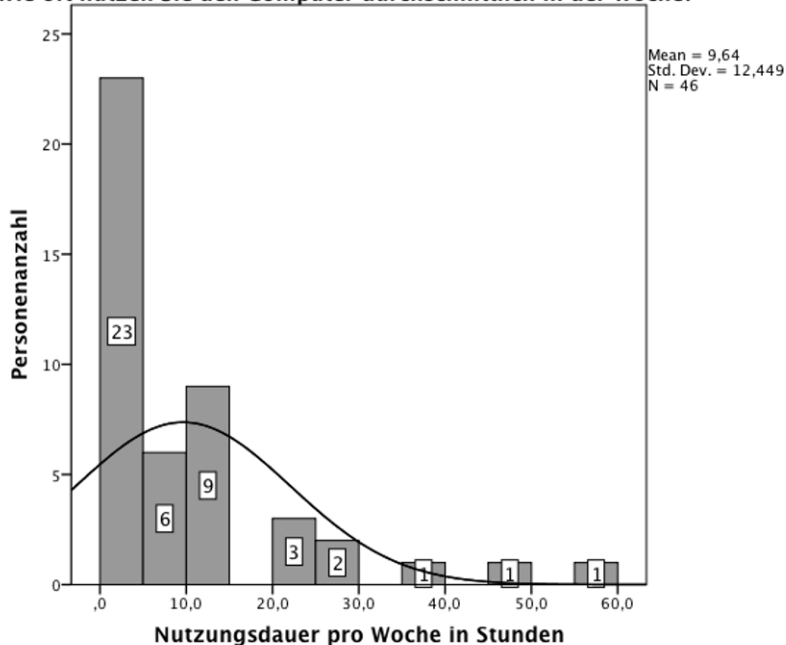
Wie oft nutzen Sie den Computer durchschnittlich in der Woche?

Abbildung 34: Balkendiagramm durchschnittliche Nutzungsdauer des Computers pro Woche

Die Nutzungsdauer liegt bei durchschnittlich 9,64 Stunden in der Woche (SD=12,44). Die Verteilung zeigt allerdings, dass ein Großteil den Computer unter 15 Stunden wöchentlich nutzt (38 Personen, 76%), wovon sieben Personen nie den Computer nutzen. Acht Personen (16%) nutzen den Computer über 15 Stunden wöchentlich, drei Personen (6%) davon über 35 Stunden.

34 Personen (68%) haben angegeben, Zugang zum Internet zu haben. Hinsichtlich der Nutzung des Internets machten die Probanden und Probandinnen folgende Angaben:

	nie	weniger als 1x / Woche	mind. 1x / Woche	mind. 1x / Tag	Fehlend	Gesamt
Surfen	20	4	7	18	1	50
	40%	8%	14%	36%	2%	100,0%
Nachricht ansetzen	25	6	8	9	2	50
	50%	12%	16%	18%	96%	100,0%
Emailnutzung	27	5	7	10	1	50
	54%	10%	14%	20%	2%	100,0%
Einkaufen	39	5	3	1	2	50
	78%	10%	6%	2%	4%	100,0%
Telefonieren	38	3	3	3	3	50
	76%	6%	6%	6%	6%	100,0%
Musiknutzung	18	4	8	18	2	50
	36%	8%	16%	36%	4%	100,0%
Videonutzung	26	9	7	5	3	50
	52%	18%	14%	10%	6%	100,0%
Fernsehen	37	3	1	6	3	50
	74%	6%	2%	12%	6%	100,0%
Spiele spielen	25	5	9	7	4	50
	50%	10%	18%	14%	8%	100,0%
Nutzung von Sozialen Netzwerken	21	2	6	18	3	50
	42%	4%	12%	36%	6%	100,0%

Tabelle 33: Internetnutzung

20 Personen (40%) surfen nie im Internet, 18 Personen (36%) mindestens einmal am Tag. Sieben Personen (14%) nutzen das Internet zum Surfen mindestens einmal, vier Personen (8%) weniger als einmal pro Woche (fehlend n=1). 25 Personen (50%) sehen nie Nachrichten über das Internet, täglich tun dies neun Personen (18%). Mindestens einmal pro Woche sehen acht Personen (16%) Nachrichten über das Internet, weniger als einmal pro Woche sechs Personen (12%) (fehlend n=2). Emails werden von 27 Personen (54%) nie genutzt, von zehn Personen (20%) täglich. Mindestens einmal pro Woche nutzen sieben Personen (14%) Emailverkehr, fünf Personen (10%) weniger als einmal pro Woche (fehlend n=1). 39 Personen (78%) kaufen nie über das Internet ein, 5 Personen (10%) weniger als einmal pro Woche und vier Personen (8%) mindestens einmal pro Woche. Eine Person (2%) gibt an, mindestens täglich im Internet einzukaufen. 38 Personen (78%) geben an, das Internet nie zum Telefonieren zu nutzen, jeweils drei Personen (6%) tun das weniger als einmal pro Woche bzw. mind.

Einmal pro Woche bzw. mindestens täglich (fehlend n=3). Musik nutzen 18 Personen (36%) täglich und genauso viele Personen nie via Internet. Vier Personen (8%) beschäftigen sich weniger als einmal pro Woche und acht Personen (16%) mindestens wöchentlich mit der Musiknutzung (fehlend n= 2). Videos sehen sich 26 Personen (52%) nie und nur fünf Personen (10%) täglich im Internet an. Neun Personen (18%) nutzen Videos weniger als einmal pro Woche und 7 Personen (14%) mindestens einmal pro Woche (fehlend n= 3). Zum Fernsehen nutzen 37 Personen (74%) nie, sechs Personen (12%) täglich, drei Personen (6%) weniger als einmal pro Woche und eine Person (2%) mindestens einmal pro Woche das Internet (fehlend n= 3). Über das Internet spielen 25 Personen (50%) nie und sieben Personen (14%) täglich spiele. Weniger als einmal in der Woche tun dies fünf Personen (10%) und mehr als einmal in der Woche neun Personen (18%) (fehlend n= 4). Für Soziale Netzwerke nutzen 21 Personen (42%) nie und 18 Personen (36%) täglich das Internet. Weniger als einmal pro Woche nutzen zwei Personen (10%) und mindestens wöchentlich sechs Personen (12%) soziale Netzwerke (fehlend n=3).

Die Angaben zeigen, dass ähnlich viele Personen das Internet zum Surfen, zur Musiknutzung sowie für soziale Netzwerke nie nutzen wie Personen, die es täglich dafür nutzen. Die größte Diskrepanz zeigt sich bei der Nutzung des Internets zum Einkaufen. Dass 39 Personen (78%) das Internet nie zum Einkaufen nutzten, ist vermutlich weniger in der Computererfahrung, sondern stärker in deren Kaufkraft begründet (vgl. Kap. 2.3.3 zu funktionalen Analphabeten und Analphabetinnen und Erwerbstätigkeit).

Weitere Angaben zur Nutzung des Computers, die von mehreren Personen genannt wurden sind Lernen (inkl. der Nutzung von Lernprogrammen) (n=10), Bewerbung bzw. Stellensuche (n=6) und Chatten (n=6)⁹⁹.

Computerkompetenz

Die Ergebnisse des Fragebogens zur Computerkompetenz zeigen, dass die Mehrzahl der Personen die grundlegenden Fähigkeiten zur Bedienung von otu.lea beherrschte: Sie bringen das Wissen über die Funktionen von Hochsteltaste, Cursorplatzierung durch Betätigung der linken Maustaste sowie Eingabe einer Internetadresse in ein Adressfeld im Browser mit: 90% (45 Personen) der Befragten haben die richtigen Felder zur Eingabe der Internetadresse gewählt, sechs Prozent (3 Personen) haben das falsche Eingabefeld gewählt und vier Prozent (2 Personen) haben keine Angabe gemacht. Die Frage zur Eingabe in ein Textfeld (Cursorplatzierung und Eintippen von Buchstaben) haben 92% (46

⁹⁹ Die Häufigkeit der Nutzung wurde hierbei nicht erfragt.

Personen) richtig, sechs Prozent (3 Personen) falsch und zwei Prozent (eine Person) mit „weiß nicht“ beantwortet. Das Wissen über die Hochstelltaste zur Nutzung von Großschreibung hatten zur Zeit der Befragung 70% (35 Personen), wohingegen 30% (15 Personen) die Frage falsch beantworteten.¹⁰⁰

Da die meisten Personen die Fragen richtig beantwortet haben, zeigt das Antwortverhalten eine sehr geringe Varianz. Weitere inferenzstatistische Untersuchungen hinsichtlich eines möglichen Effektes der Computerkompetenz werden daher nicht durchgeführt.

Die bisher dargestellten Ergebnisse werden mitunter in den folgenden Kapiteln für die Beantwortung der Fragestellungen aufgegriffen.

8.3.3 *Ergebnisdarstellung zu den Usability-Heuristiken*

Bearbeitung der Frage Fl1a „Werden die als für funktionale Analphabeten und Analphabetinnen relevant identifizierten Heuristiken in otu.lea berücksichtigt?“

Um die Frage Fl1a „Werden die als für funktionale Analphabeten und Analphabetinnen relevant identifizierten Heuristiken in otu.lea berücksichtigt?“ zu beantworten, werden je nach Heuristik die relevanten Datenquellen (Computererfahrung, Performanz, UEQ, Interviews und Videos) herangezogen und ausgewertet (vgl. Kap. 8.1.2.1 zur Operationalisierung).

Erlernbarkeit

Anzeigend für die Erlernbarkeit ist, wie sich die eingeforderte Unterstützung in Bezug zur Gesamtbearbeitungszeit verhält (die Gesamtbearbeitungszeit ist in den Tabellen 8 und 9 im Anhang aufgeführt; Angaben zur Unterstützung s. Tabelle 27 in Kap. 8.3.2.2). Die Gesamtbearbeitungszeit wird über die Videos ermittelt. Wird zu Beginn der Nutzung von otu.lea mehr Unterstützung eingefordert als am Ende, ist dies ein Hinweis darauf, dass otu.lea „erlernbar“ ist. Für die Bewertung der Erlernbarkeit wurden die Codings für die Kategorien „eingeforderte inhaltli-

¹⁰⁰ Die Angaben in dem Fragebogen wurden mit den Videobeobachtungen hinsichtlich des Einsatzes der Hochstelltaste sowie der Cursorplatzierung verglichen. Drei Probanden/Probandinnen (A3, U3, AB3) wurden im Laufe des Usability-Tests die Notwendigkeit der Cursorplatzierung zur Eingabe von Buchstaben in einem Textfeld erläutert, zwei Probanden/Probandinnen (AD3, AN3) wurden sowohl Hochstelltaste als auch Cursorplatzierung erläutert, einer Person (M3) wurde die Hochstelltaste erläutert. Damit ist bei diesen Personen davon auszugehen, dass sie vor dem Usability-Test diese Fähigkeiten bzw. das Wissen über diese Funktionen nicht hatten. Diese Befunde weichen von denen im Fragebogen ab. Werden die Beobachtungen berücksichtigt, ergibt sich folgende Verteilung: Das Wissen über die Eingabe in ein Textfeld hätten voraussichtlich 42 Personen, über die Funktion der Hochstelltaste 34 Personen.

che Unterstützung“ und „eingeforderte funktionale Unterstützung“ sowie die Gesamtbearbeitungszeit verwendet. Aus den Zeitpunkten der Unterstützung wird der Mittelwert berechnet, ebenso wird der Mittelwert der Gesamtbearbeitungszeit berechnet. Befindet sich der durchschnittliche Wert für die eingeforderte Unterstützung innerhalb des unteren Bereichs der Skala der durchschnittlichen Gesamtbearbeitungszeit, ist die Erlernbarkeit ausgeprägt. Hingegen ist die Erlernbarkeit wenig bzw. nicht vorhanden, wenn sich der Mittelwert des Zeitpunkts der eingeforderten Unterstützung in der zweiten Hälfte der Skala der Bearbeitungszeit befindet. Videos, aus denen die Gesamtbearbeitungszeit nicht hervorgeht, werden nicht in die Auswertung einbezogen.

Über alle Probanden und Probandinnen hinweg beträgt der Mittelwert der Gesamtbearbeitungszeit 36 Minuten und zwei Sekunden ($SD=00:14:26$, Minimum= $00:11:00$, Maximum= $01:22:18$) (für eine ausführliche Darstellung siehe Tabellen 8 und 9 im Anhang). Das Video AH3 wird aufgrund einer zu kurzen Aufnahme nicht in die Berechnung einbezogen. Der Zeitpunkt der Unterstützung liegt durchschnittlich bei 16 Minuten und 45 Sekunden ($SD=00:16:01$, Minimum= $00:00:00$, Maximum= $01:06:57$) ($n=31$). Personen, die keine Unterstützung einforderten, werden in den Berechnungen nicht berücksichtigt (17 haben keine inhaltliche und 28 keine funktionale Unterstützung eingefordert). Somit liegt der Zeitpunkt, zu dem im Durchschnitt Unterstützung eingefordert wurde, in der unteren Hälfte (wobei 16,46 Minuten 45,67% von $00:36:02$ Minuten entsprechen). Die Konfidenzintervalle für beide Variablen zeigen keine Überschneidungen (s. Abbildung 35). Daher ist von einem signifikanten Unterschied auszugehen.

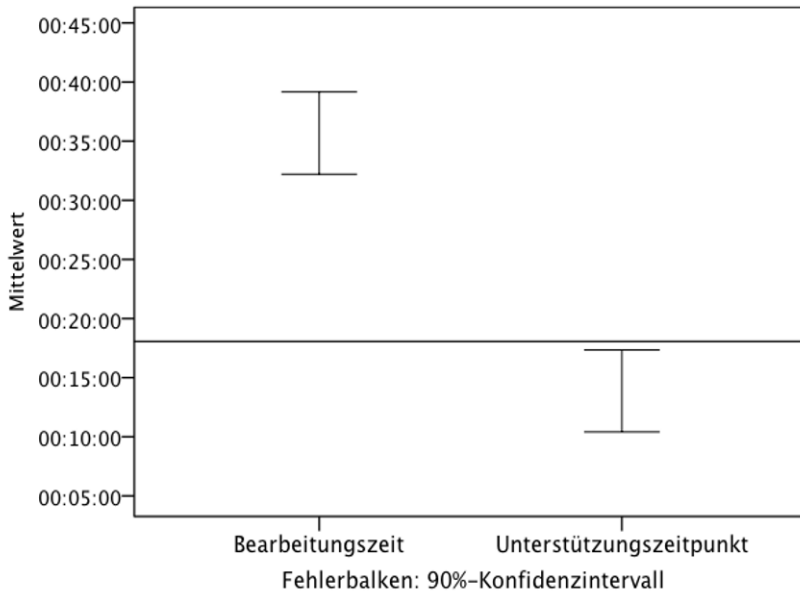


Abbildung 35: Durchschnittliche Gesamtbearbeitungszeit im Verhältnis zur durchschnittlichen eingeforderten Unterstützung

Um nur die Probanden und Probandinnen mit wenig Computererfahrung zu berücksichtigen, wird über die Items zur Internetnutzung der Personenwert „Computererfahrung“ errechnet. Entsprechen der Antwortskala der Items werden Personen, die nie oder weniger als einmal wöchentlich das Internet nutzen der Kategorie „wenig Erfahrung“ zugeordnet. Personen, dessen errechneter Durchschnittswert in den Bereichen mindestens wöchentlich oder mindestens täglich zu verorten ist, werden der Kategorie „viel Erfahrung“ zugeordnet. Auf die Kategorie „wenig Erfahrung“ fallen 27 und auf die Kategorie „viel Erfahrung“ 22 Probanden und Probandinnen. Eine Person hat die Fragen nicht beantwortet.

Werden nur die Personen mit wenig Erfahrung in die Berechnungen einbezogen, ergeben sich folgende Werte: Der Mittelwert der Personen mit wenig Computererfahrung für die Gesamtbearbeitungszeit beträgt 41 Minuten und 6 Sekunden (SD=00:16:43, Minimum=00:17:24, Maximum=01:22:18) (n=25). Auch hier werden wieder diejenigen, die keine Unterstützung einforderten, nicht

in die Berechnung einbezogen ($n=2$)¹⁰¹. Der Mittelwert des Unterstützungszeitpunktes liegt bei 17 Minuten und 41 Sekunden ($SD=00:16:16$, Minimum=00:00:00, Maximum=01:06:57) und befindet somit ebenfalls in der ersten Hälfte der Gesamtbearbeitungszeit der Personen mit wenig Computererfahrung. Die prozentuale Berechnung zeigt, dass der durchschnittliche Zeitpunkt der Unterstützung bei 42,4% liegt. Damit liegt der Mittelwert der eingeforderten Unterstützung geringfügig niedriger bei Personen mit wenig Computererfahrung im Vergleich zur Gesamtstichprobe. Auch hier überschneiden sich die Konfidenzintervalle nicht (s. Abbildung 36).

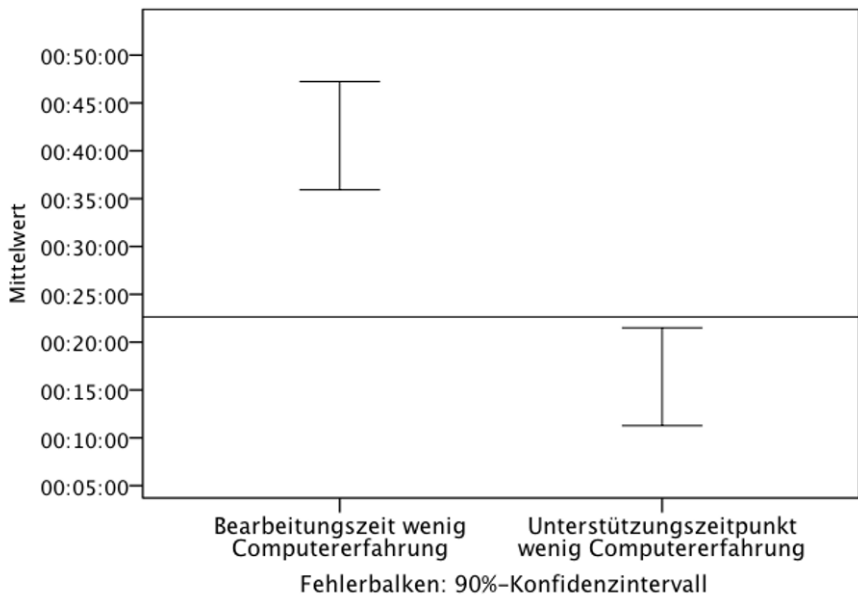


Abbildung 36: Vgl. Bearbeitungszeit und Unterstützungszeitpunkt (wenig Computererfahrung)

Somit ist für beide Berechnungen (durchschnittlicher Unterstützungszeitpunkt für Personen mit viel sowie für Personen mit wenig Computererfahrung) von einem signifikanten Unterschied auszugehen.

¹⁰¹ Für eine Person (AM3) liegen keine Ergebnisse zur Computererfahrung vor. Eine Person (R3) hat sich tatsächlich ohne weitere Unterstützung in der Umgebung zurechtgefunden trotz niedriger Computererfahrung.

Zu berücksichtigen ist, dass 32 von 50 Probanden und Probandinnen zwei Dimensionen bearbeitet haben und in den Dimensionen unterschiedliche Aufgabenformate vorkommen. Wurde die Aufgabenbearbeitung einer Dimension neu begonnen, ist der Unterstützungsbedarf aufgrund neuer Formate (z. B. die Eingabe von Buchstaben in ein Textfeld) mit hoher Wahrscheinlichkeit höher als zum Ende der Bearbeitung der vorherigen Dimension. Würden die Dimensionen einzeln in die Berechnung einbezogen, wäre der durchschnittliche Zeitpunkt der Unterstützung (auch relativ zur Bearbeitungszeit einer Dimension) vermutlich früher als die oben dargestellten Ergebnisse zeigen. Des Weiteren ist anzumerken, dass Personen, die keine Unterstützung eingefordert haben, in den Berechnungen nicht berücksichtigt wurden. D.h. in der ersten Berechnung zur Gesamtbearbeitungszeit und eingeforderten Unterstützung für alle Probanden und Probandinnen sind nur 31 Personen (von 50) berücksichtigt. Für die Berechnungen für Personen mit wenig Computererfahrung wurden 25 Personen einbezogen. Werden diese beiden Teilstichprobengrößen in Beziehung gesetzt wird deutlich, dass hauptsächlich Personen mit wenig Computererfahrung Unterstützung einforderten. Würden auch die Personen in die Berechnungen einbezogen, die keine Unterstützung einforderten, wäre der durchschnittliche Zeitpunkt der eingeforderten Unterstützung vermutlich früher als in den eben durchgeführten Berechnungen.

Hinsichtlich der Aussagekraft der Ergebnisse ist zu erwähnen, dass der Effekt der Unterstützung abhängig von der Qualität der Unterstützung sein kann. Beispielsweise kann die Unterstützung eines Betreuers/einer Betreuerin in einem Maße gut verständlich und umfassend gewesen sein, so dass anschließend nur wenig Unterstützung notwendig war. Hingegen kann die Unterstützung eines anderen Betreuers/einer anderen Betreuerin weniger umfassend oder auch verständlich ausgefallen sein, was möglicherweise einen hohen Unterstützungsbedarf auch im weiteren Verlauf der Testung zur Folge hat. Die Videos zeigen dahingehend allerdings keine auffälligen Diskrepanzen zwischen den betreuten Personen.

Fazit zur Erlernbarkeit

Für funktionale Analphabeten und Analphabetinnen scheint insbesondere eine intensive Unterstützung zu Beginn notwendig zu sein. Zu Beginn ist eine umfassende Unterstützung durch das System durch wiederholte Hinweise auf die Hilfefunktion sowie erläuternde Pop Ups, die z. B. erscheinen, wenn über einen bestimmten Zeitraum (ca. fünf Minuten) keine Aktion ausgeführt wurde bzw. hinsichtlich der Codeeingabe mehrfach der falsche Code eingegeben wurde. Da sich die Ergebnisse zur Erlernbarkeit ausschließlich auf die Bearbeitungszeit und die Unterstützung durch betreuende Person beziehen, dienen die Ergebnisse nur bedingt einer generalisierbaren Aussage zur Erlernbarkeit. Unter Berücksich-

tigung der Operationalisierung der Erlernbarkeit ist diese Heuristik allerdings als erfüllt zu bewerten.

Selbstbeschreibungsfähigkeit

Zur Beobachtung und Beurteilung der Selbstbeschreibungsfähigkeit sind der Effekt der Hilfefunktionsnutzung sowie die eingeforderte Unterstützung anzeigend. Die Hilfefunktion wird für die gesamte Testumgebung beobachtet. Da sich die Login-Sequenz stark von dem Aufgabenbereich unterscheidet (die Seiten der Login-Sequenz werden nicht iterativ aufgebaut; des Weiteren besitzt der Login-Bereich dadurch, dass der Login zwingend erforderlich für die Nutzung von otu.lea ist, eine besondere Relevanz) werden für die vertiefende Auswertung mit Hilfe der Videos der Aufgabenbereich und die Login-Sequenz getrennt analysiert.

Darstellung der Beobachtungen zum Aufgabenbereich

Die Hilfefunktion ist auf jeder Seite abrufbar und beinhaltet teilweise kontext-sensitive Videos. Für den Prescreen war zum Zeitpunkt der Studie kein kontextsensitives Hilfevideo implementiert und es wurde das Video, in dem die Seitenstruktur erläutert wird, eingesetzt.

	Häufigkeit	Prozent
Gültig		
0	42	84,0
1	4	8,0
3	3	6,0
4	1	2,0
Gesamtsumme	50	100,0

Tabelle 34: Nutzung der Hilfefunktion

Die Hilfefunktion wurde insgesamt von acht Personen genutzt, wobei vier Personen die Funktion einmal, drei Personen dreimal und eine Person viermal aktivierten.

Der Proband/Die Probandin L3 aktiviert erstmals die Hilfefunktion bei einer Korrekturaufgabe¹⁰². Nach der Erläuterung der Lautsprecher- und Vorlesebuttons beendet der Proband/die Probandin die Hilfefunktion und kehrt zur Aufgabenbearbeitung zurück. Die Person klickt auf den Weiter-Button, um zur nächsten Aufgabe zu gelangen, ohne die Aufgabe zu bearbeiten. Ungefähr 20 Minuten später aktiviert Proband/Probandin L3 erneut die Hilfefunktion. Zu dem Zeitpunkt befindet er/sie sich in der Dimension Schreiben auf mittlerem Niveau. Im Aufgabenbereich (auf der linken Seite) sind bereits der Einführungstext und ein

¹⁰² Die Hilfefunktion besteht auf dieser Seite aus der formatübergreifenden Erläuterung der Seitenstruktur und Funktionen.

Bild, im Arbeitsbereich sind noch keine Informationen angezeigt. Die Person sieht sich das Video bis zum Ende an. Kurz vor Ende versucht sie, das Video zu beenden, indem sie im ausgegrauten¹⁰³ Bereich hinter dem Video auf den Startbutton klickt, der normalerweise – wenn das Hilfevideo nicht läuft – die Aufgabe aufruft. Nach dem Ende des Videos klickt die Person weiterhin auf den Weiter-Button im ausgegrauten Bereich und startet das Video anschließend neu. Nach ca. 20 Sekunden aktiviert sie den Weiter-Button unter dem Hilfevideo und verlässt damit die Hilfefunktion. Anschließend wird der Startbutton aktiviert und die Person beginnt mit der Aufgabenbearbeitung. Das dritte Mal wird von Proband/Probandin L3 die Hilfefunktion bei derselben Aufgabe aktiviert, nachdem sie auf einen Lautsprecherbutton eines Textfeldes geklickt hat, dessen Audiodatei fehlt (ein technischer Fehler). Bereits nach ein paar Sekunden versucht die Person, das Hilfevideo zu beenden, indem sie in dem ausgegrauten Bereich den Weiter-Button anklickt. Die Person sieht sich das Video bis zum Ende an, klickt erneut auf den Weiter-Button im ausgegrauten Bereich und schließlich auf den Weiter-Button unter dem Video, so dass sie zum Aufgabenbereich zurückkehrt. Nach mehrmaligem Klicken auf den Lautsprecher, beendet sie die Aufgabenbearbeitung und beginnt die nachfolgende Aufgabe.

Person S3 aktiviert die Hilfefunktion auf der Seite der Niveaue Auswahl, nachdem sie zwischenzeitlich zur Dimensionsauswahl zurückkehrte. Die Person sieht sich das Video allerdings nur bis zum ersten Drittel an und kehrt zur Seite der Niveaue Wahl zurück. Anschließend fordert sie Unterstützung von der betreuenden Person ein. Die Proband/Die Probandin wird gefragt, ob er/sie bereits ein Niveau ausgewählt hätte. Der Proband/Die Probandin gibt an, bereits die Schaltfläche zur Niveaue Wahl aktiviert zu haben, es sei aber nichts passiert, er/sie aktiviert währenddessen die Schaltfläche des Niveaus „einfach“ und gelangt in den Aufgabenbereich.

Die Person V3 aktiviert die Hilfefunktion, nachdem zum ersten Mal ein Prescreen erscheint. Zum Prescreen erscheint ebenfalls das Video zur Erläuterung der Seitenstruktur und der grundsätzlichen Funktionen. Die Person sieht sich das Video bis zum Ende an und aktiviert anschließend den Soundbutton des Prescreens, wodurch die Person schließlich zur Aufgabenbearbeitung gelangt. Der Proband/Die Probandin W3 aktiviert die Hilfefunktion während das Einführungstutorial und in diesem auf die Hilfefunktion hingewiesen wird¹⁰⁴. Erneut wird die Hilfefunktion nach erstmaligem Erscheinen des Prescreens aktiviert, als

¹⁰³ „Ausgegraut“ ist der Bereich hinter dem Hilfevideo. Das Video wird in einem separaten Fenster abgespielt, das kleiner als das eigentlich Interface ist. Hinter dem Video ist aber dennoch die Seitenstruktur zu sehen, jedoch ist in dem Bereich keine Aktivität möglich, wenn das Video noch geöffnet ist.

¹⁰⁴ In dem Hilfevideo werden die Seite des Einführungstutorials und die Schaltflächen erläutert.

lediglich zwei Grafiken und der Soundbutton auf dem Prescreen zu sehen sind. Die Person sieht sich das Video bis zum Ende an und aktiviert mit Unterstützung des Betreuers/der Betreuerin den Soundbutton. Nachdem sich der Text angehört wurde und der Weiter-Button erschienen ist, aktiviert die Person erneut die Hilfefunktion und sieht sie sich fast bis zur Hälfte an. Anschließend klickt sie den Weiter-Button und gelangt zur ersten Aufgabe. Das Programm stürzt kurze Zeit später ab und die Person beginnt erneut. Die Hilfefunktion wird erneut an der gleichen Stelle beim Prescreen genutzt, nachdem die Audiodatei bereits abgespielt und der Weiter-Button angezeigt wurde. Die Hilfefunktion wird wieder nach ca. der Hälfte beendet und die Person klickt auf den Weiter-Button.

Wie auch Proband/Probandin W3 bei der erstmaligen Aktivierung der Hilfefunktion, ruft Person AE3 ebenfalls die Hilfefunktion während des Einführungstutorials auf, unmittelbar nach dem Hinweis auf die Hilfefunktion. Mit Unterstützung des Betreuers/der Betreuerin kehrt die Person zum Einführungstutorial zurück. Das zweite Mal wird die Hilfefunktion nach erstmaligem Erscheinen des Prescreens aufgerufen. Auf dem Prescreen sind bereits die Einführungstexte und der Weiter-Button abgebildet. Die Person sieht sich das Video bis ca. zur Hälfte an und klickt, nachdem sie auch versucht hat, im Video Schaltflächen zu aktivieren, auf „Weiter“ und kehrt zum Prescreen zurück. Dort betätigt sie die Soundbuttons und schließlich die Schaltfläche, um zur Aufgabenbearbeitung zu gelangen. Ein letztes Mal ruft der Proband/die Probandin die Hilfefunktion im Aufgabenbereich auf, nachdem auf der linken Seite der Einführungstext und ein Stimulus zur Aufgabenbearbeitung erschienen sind. Nach wenigen Sekunden wird das Video unterbrochen und die Person ruft unmittelbar über den Startbutton die zu bearbeitende Aufgabe auf.

Der Proband/Die Probandin AM3 befindet sich zum Zeitpunkt der Hilfefunktionsnutzung in der Dimension Lesen, hat bereits einige Aufgaben bearbeitet und gelangt zu einer neuen Aufgabenseite, auf der nur der Lautsprecherbutton und die Hilfefunktion zu sehen sind. Die Person verweilt ein paar Sekunden ohne Aktion. Die betreuende Person macht den Proband/die Probandin auf die Hilfefunktion aufmerksam, woraufhin der Proband/die Probandin die Hilfe aktiviert. Die Hilfefunktion zur Erläuterung der Seiten- und Funktionserläuterung wird bis zum Ende angesehen und versucht, die Schaltfläche im ausgegrauten Bereich zu aktivieren. Mit Hilfe des Betreuers/der Betreuerin (durch das Zeigen auf die zu aktivierende Schaltfläche, um die Hilfefunktion verlassen zu können) kehrt der Proband/die Probandin wieder zur Aufgabenseite zurück. Der Proband/Die Probandin aktiviert nach kurzer Verweildauer auf der Seite erneut die Hilfefunktion. Durch die unaufgeforderte Unterstützung des Betreuers/der Betreuerin aktiviert sie nach Beendigung schließlich den Soundbutton und fährt mit der Aufgabenbearbeitung fort. Ein weiteres Mal wird die Hilfefunktion nach

Erscheinen des Prescreens aufgerufen. Die Person sieht sich erneut die Hilfefunktion zur Seiten- und Funktionserläuterung bis zum Ende an und klickt anschließend auf den Soundbutton auf dem Prescreen. Auch beim dritten Mal der Hilfenutzung befindet sich die Person im Bereich des Prescreens. Wieder sieht sie sich das Video bis zum Ende an und betätigt anschließend den Soundbutton. Die Person AS3 hat das mittlere Niveau im Bereich Schreiben gewählt und aktiviert die Hilfefunktion, nachdem der Einführungstext auf der linken Seite des Aufgabenbereichs vorgelesen und angezeigt wurde. Sie beendet das Hilfevideo zur Seiten- und Funktionserläuterung nach ca. der Hälfte – nachdem Sound- und Lesesymbol erläutert wurden – klickt unmittelbar danach das Lesesymbol an und beginnt mit der Aufgabenbearbeitung.

Der Proband/Die Probandin Q3 ruft die Hilfefunktion während des Einführungstutorials auf, nachdem das Programm nach der Bearbeitung einer Aufgabe abgestürzt ist. Der Proband/Die Probandin kommentiert das Erscheinen des Hilfevideos mit einem „Oh“, was darauf hindeutet, dass die Person das Abspielen eines Videos nicht erwartet hat. Sie sieht sich das Video bis zum Ende an und versucht anschließend über eine Schaltfläche im ausgegrauten Bereich zurückzukehren. Der Betreuer/Die Betreuerin macht den Probanden/die Probandin schließlich auf den Weiter-Button aufmerksam, der das Verlassen der Hilfefunktion ermöglicht.

Vorerst ist für die Selbstbeschreibungsfähigkeit hinsichtlich der Hilfefunktion festzustellen, dass der Prescreen und das Beenden der Hilfevideos nicht ausreichend transparent und selbsterklärend dargestellt sind.

Darstellung der Beobachtungen zur Login-Sequenz

Die Login-Sequenz erstreckt sich von der Seite, auf der die teilnehmende Person gefragt wird, ob sie bereits einen Test absolviert hat bis hin zur Seite, auf der die Eingabe des (neuen) Codes gefordert wird. Zur Beurteilung der Selbstbeschreibungsfähigkeit wird für diesen Bereich zusätzlich zur Hilfefunktionsnutzung mit Hilfe der Videos beobachtet, wie oft die Probandinnen und Probanden Unterstützung erhalten und welche Herausforderungen für sie dabei bestehen.

Login	n
Login mit Unterstützung	32
Login ohne Unterstützung	5
kein Login beobachtbar	13
Gesamt	50

Tabelle 35: Unterstützung beim Einloggen

Die Beobachtungen zur Login-Sequenz zeigen, dass die Mehrheit der Probanden und Probandinnen zum Einloggen Erläuterungen durch die Betreuer und Betreu-

erinnen bedurften. Wie die Tabelle 35 zeigt, erhielten 32 Personen (64%) Unterstützung, lediglich fünf Personen meldeten sich ohne Unterstützung an. Von 13 Personen war der Login nicht beobachtbar, entweder weil die Videoaufzeichnung zu spät begann oder kein Video vorhanden ist. Zu beobachten ist, dass die Probanden und Probandinnen nicht wissen, welchen Code sie eintragen sollen. Primär zeigen die Videos, dass die Personen zu Beginn nicht den Lautsprecher anklicken, um sich den Text vorlesen zu lassen, sondern sich bemühen, den Text selbst zu lesen. Auch geben Probanden und Probandinnen an, bereits einen Test absolviert zu haben. Einige Probanden und Probandinnen gehen davon aus, dass sie ihren Namen eingeben oder sich etwas ausdenken können.

Allerdings hat hinsichtlich der Aufgabenbearbeitung die Mehrheit der Probanden und Probandinnen keine Unterstützung eingefordert (s. Tabelle 27 im Kap. Allgemeine Ergebnisse der Videos). Auch unaufgefordert wurde der Großteil (funktionale und inhaltliche Unterstützung zusammengerechnet) nicht unterstützt. Das deutet auf eine recht ausgeprägte Selbstbeschreibungsfähigkeit des Aufgabenbereichs von otu.lead hin.

Für die Nutzung der Hilfefunktion ist der häufigste Nutzungszeitpunkt (fünf von 17) der Prescreen. Die Nutzung der Hilfe auf dem Prescreen ist vermutlich darin begründet, dass die Seitenstruktur vorerst unbekannt und durch kein visuell zweigeteiltes Interface recht leer erscheint im Vergleich zur Aufgabenseite. Während des Einführungstutorials wird dreimal die Hilfefunktion aktiviert. Das Aufrufen der Hilfefunktion während des Einführungstutorials ist wahrscheinlich darin begründet, dass während des Tutorials auf die Hilfefunktion hingewiesen wird und die Probanden und Probandinnen diesen Hinweis als Aufforderung zur Nutzung der Hilfefunktion verstehen. Bei Person L3 ist verwunderlich, dass sie die Hilfefunktion genutzt hat, obwohl sie bereits ähnliche Aufgaben mit der gleichen Seitenstruktur und Funktionen bearbeitet hat. Erklären lässt sich der Unterstützungsbedarf vermutlich dadurch, dass es sich bei der Aufgabensequenz um Teilaufgaben handelte. Bei Teilaufgaben sind die Einführung und die Stimuli meist identisch. Vermutlich hat der Proband/die Probandin den Eindruck, wieder zurückgekehrt zu sein, da die Seite der nachfolgenden Teilaufgabe mit der Seite der vorherigen Teilaufgabe identisch ist.

Das Beenden der Hilfefunktion scheint nicht transparent und selbsterklärend zu sein. Vier Personen (L3, AE3, AM3 und Q3) erhalten Unterstützung, um das Video zu beenden und in die Testumgebung zurückzukehren. Drei Personen (L3, AM3, Q3) versuchen, Schaltflächen in dem ausgegrauten Bereich zu aktivieren. Vermutlich ist die Ursache, dass der Start- bzw. Weiter-Button, auf dem in dem Video hingewiesen wird, während des Videos noch zusehen ist. So gehen Probanden und Probandinnen vermutlich davon aus, dass sie diesen Button auch nach dem Hilfevideo aktivieren müssen. Das Hilfevideo auf der Seite des Ein-

führungstutorials weist ebenfalls auf den Startbutton hin, so dass die Probanden und Probandinnen anscheinend auch hier davon ausgehen, den Button, der im Video erläutert wird, unmittelbar anklicken zu müssen (Proband/Probandin AE3). Vermutlich ist die Transferleistung, dass es sich bei den im Video erläuterten Schaltflächen nicht um die Schaltflächen zur Beendigung des Videos handelt, durch mangelnde Erklärung seitens der Versuchsleitenden eingeschränkt. Dass die Hilfefunktion von den gleichen Probanden und Probandinnen an gleichen Stellen bzw. Stellen mit gleichem Aufbau genutzt wird ist ein Hinweis darauf, dass es sinnvoll ist, permanent Hinweise zur Bearbeitung und insbesondere zur Reihenfolge der Aktionen zu geben. Eine Person nutzt die Hilfefunktion, weil sie vermutlich davon ausgeht, dass keine neue Aufgabe erschienen ist, obwohl sie auf „Weiter“ geklickt hat. Sinnvoll erscheint daher, den Wechsel zwischen Teilaufgaben visuell deutlich zu gestalten, so dass die Teilnehmenden sehen, dass es sich um eine neue Aufgabe handelt.

Die Hilfefunktion führt in zwei Fällen (S3 und W3) nicht dazu, dass die Probanden und Probandinnen nach dem Ansehen des Videos selbstständig weiterarbeiten. Die Probanden und Probandinnen erhalten für weitere Aktionen die Unterstützung des Betreuers/der Betreuerin. Im Fall von S3 liegt es vermutlich daran, dass die Person durch den im Vorfeld erzeugten vermeintlichen Eindruck, das Klicken auf die Schaltflächen der Niveauewahl scheine nicht zu funktionieren, auch den Inhalt des Hilfevideos nicht als sinnvoll empfindet. Denn in dem Video wird das erläutert, wovon sie ausgeht, dass es nicht funktioniert. Im Fall von W3 ist die Hilfefunktion für diese Person vermutlich nicht hilfreich bzw. unverständlich oder der Proband/die Probandin sieht sich den Inhalt des Videos nicht aufmerksam genug an.

Insgesamt wird die Hilfefunktion von nur fünf Personen genutzt. Demgegenüber stehen 17 Personen, die mindestens einmal inhaltliche und 28 Personen, die mindestens einmal funktionale Unterstützung einforderten.

Dass der Großteil Unterstützung beim Einloggen benötigte scheint darin begründet zu sein, dass die Probanden und Probandinnen entweder nicht beantworten können, ob sie bereits *otu.lea* genutzt haben (vermutlich aufgrund dessen, weil sie die Frage nicht verstehen oder andere Programme nutzen und nicht zwischen diesen und *otu.lea* unterscheiden oder seitens des Programms nicht deutlich auf den einzugebenden Code hingewiesen wird). Ein weiterer Grund könnte darin liegen, dass die Probanden und Probandinnen zu Beginn nervös waren und daher die Konzentrationsfähigkeit nicht ausgeprägt genug war, um die Fragen und Anweisungen korrekt zu interpretieren. Funktionale Analphabeten und Analphabetinnen scheinen einerseits Funktionen, die bereits erläutert oder auch genutzt wurden, zu vergessen. Andererseits haben sie Hemmschwellen, explorativ vorzugehen.

Fraglich ist zudem, ob funktionale Analphabeten und Analphabetinnen es überhaupt gewöhnt sind, eine derartige Hilfefunktion in Anspruch zu nehmen. Zum einen kann die Nutzung einer solchen Hilfefunktion unbekannt sein, da funktionale Analphabeten und Analphabetinnen tendenziell weniger mit dem Computer arbeiten als literalisierte Personen und somit wenig Erfahrung in der Nutzung von Computern auch damit auch der Hilfefunktion haben. Zum anderen haben funktionale Analphabeten und Analphabetinnen, die Computer nutzen, womöglich die Erfahrung gemacht, dass die Hilfefunktionen textlastig sind. Im Softwarebereich werden Hilfefunktion häufig in Form von Textmanuals angeboten, die durchsucht werden können. Für wenig literalisierte Personen ist die Nutzung einer Hilfefunktion in Form eines Textmanuals u. U. wenig hilfreich, da sie nicht in der Lage sind, sich den Text zu erschließen oder ein Wort in das Suchfeld des Manuals einzugeben. Es ist möglich, dass entsprechend dieser Erfahrung die Nutzung von Hilfefunktionen für sie mit nicht lesbaren Texten assoziiert wird. Auch diese Überlegung spricht für eine intensivere Erläuterung der Hilfefunktion.

Fazit zur Selbstbeschreibungsfähigkeit

Für das Re-Design ist festzuhalten, dass deutlicher auf die Hilfefunktion hingewiesen werden sollte. Im Prescreen sind die Cues deutlicher zu platzieren bzw. zu gestalten. Des Weiteren benötigen sie eindeutige Hinweise für die Bearbeitung der Aufgaben, ohne eine große Transferleistung bewerkstelligen zu müssen. In den Hilfevideos ist explizit zu erläutern, wie die Hilfefunktion verlassen werden kann. Zudem ist ein Hinweis einzubinden, wenn eine neue Teilaufgabe erscheint aber die Seitenstruktur, Einführungstexte und Stimuli gleich bleiben.

Für die Login-Sequenz ergeben sich folgende Modifikationsbedarfe: Die Frage, ob bereits ein Code vorhanden ist bzw. ob otu.lea bereits genutzt wurde, muss eindeutiger gestellt bzw. expliziert werden. Des Weiteren ist deutlicher darauf hinzuweisen, welcher Code eingegeben werden muss, beispielsweise durch visuelle Unterstützung und Pop-Ups, die weitere Erläuterungen anbieten.

Insgesamt ist die Selbstbeschreibungsfähigkeit für otu.lea bedingt gegeben, da sich ein kontroverses Bild zeigt. Insbesondere für den Login-Bereich ist die Selbstbeschreibungsfähigkeit von otu.lea gering und bedarf der Modifikation.

Wahrnehmungssteuerung

Die Wahrnehmungssteuerung wird durch die Überprüfung, ob ein Effekt zwischen Cues und Aufgabenbearbeitungszeit beobachtbar ist, beurteilt. Hierfür werden die Bearbeitungszeiten der 31 eingesetzten Aufgaben in einem t-Test für unabhängige Stichproben verglichen¹⁰⁵.

¹⁰⁵ Die Bedingungen für den t-Test sind erfüllt.

Der t-Test für unabhängige Stichproben für die Prüfung des Zusammenhangs zwischen dem Einsatz von Cues und der Bearbeitungszeit liefert für die erste Aufgabe im Bereich Schreiben auf einfachem Niveau (1.2.1) ein signifikantes Ergebnis ($p=.02$). Für die weiteren 30 Aufgaben ergeben sich keine signifikanten Ergebnisse.

Der Einsatz von Cues scheint somit keinen statistischen Effekt auf die Wahrnehmungssteuerung zu haben.

Ursache für den ausschließlichen Effekt bei dieser Aufgabe kann sein, dass es sich bei der Aufgabe 1.2.1 um die erste zu bearbeitende Aufgabe im Bereich Schreiben auf dem niedrigen Niveau handelt. Ggf. profitieren Personen mit weniger literalen Kompetenzen von multimedialen Hinweissignalen mehr als Personen, die bereits Aufgaben auf höheren Niveaus bearbeiten. Zudem erfordert die Bearbeitung von Leseaufgaben weniger Konstruktion und Interaktion als die Bearbeitung der Schreibaufgaben. Bei Schreibaufgaben müssen Buchstaben und Wörter selbstständig eingetippt werden. Es ist zu vermuten, dass Cues auf die Bearbeitungszeit der Schreibaufgaben keinen Effekt haben, weil die Bearbeitung der Aufgabe und somit die Kompetenz zur Bearbeitung der Aufgabe einen größeren Einfluss auf die Bearbeitungszeit haben und somit die Bearbeitungszeit vorwiegend von der Kompetenz der Probanden und Probandinnen abhängig ist. Eine inhaltsunabhängige Prüfung, ob Cues einen Effekt auf die Wahrnehmungssteuerung haben, könnte durch die getrennte Betrachtung von inhalts- und nicht inhaltsbezogenen Aktionen durchgeführt werden. Da die Bearbeitungszeit jedoch systembedingt nur aufgabenbezogen ausgegeben wurde, werden diese Berechnungen nicht durchgeführt.

Fazit zur Wahrnehmungssteuerung

Ob die Wahrnehmungssteuerung erfüllt ist, kann auf der Grundlage der vorliegenden Daten nicht beantwortet werden. Aussagekräftige Ergebnisse wären beispielsweise mit einem Eye Tracking-Experiment zu erwarten, in dem die Augenbewegungen aufgezeichnet und analysiert werden können.

Aufgabenangemessenheit

Die Aufgabenangemessenheit wird über die empfundene Aufgabenschwierigkeit und die empfundenen Herausforderungen im Umgang mit otu.lea insgesamt beurteilt. Über die empfundene Aufgabenschwierigkeit berichten die Probanden und Probandinnen heterogen, wobei der Großteil die Aufgaben als angemessen bzw. mittelmäßig anspruchsvoll ($n=24$) oder leicht ($n=13$) empfand, wie z. B. Proband/Probandin H3:

„War eigentlich leicht, ich fand da jetzt keine Schwierigkeiten bei.“ (Proband/Probandin H3)

Für einen Proband/eine Probandin waren die Aufgaben sogar irritierend leicht:

„[...] aber ich finde es halt etwas merkwürdig weil das so leicht war, man weiß halt nicht wie man sich fühlen soll“ (Proband/Probandin V3).

Allerdings ist zur Person V3 anzumerken, dass sie in beiden Dimensionen das leichte Niveau wählte.

Die Schwierigkeitssteigerung innerhalb der Niveaus wurde von einer Person bewusst wahrgenommen:

I: „Okay. Ähm, fandst du die Aufgaben denn allgemein eher schwer oder eher leicht?“

B: „Teilweise schwer und auch teilweise wieder leicht.“

I: „Okay, und welche fandst du denn jetzt schwer?“

B: „Zum Schluss die.“

I: „Ach ja, das war ja dieses Hören und dann Eintippen da, genau. Okay und das fandst du schwierig?“

B: „Ja.“ (Proband/Probandin AC3)

Ein Großteil gibt an, die Aufgaben tendenziell leicht empfunden zu haben. Vermutlich ist diese Tendenz darin begründet, dass nur wenige Personen das schwierige Niveau wählten.

I: „Ehem ok und eh wie fanden Sie die Aufgaben, eher schwer oder eher leicht?“

B: „Ja wenn du auf leicht stellst, dann ist sowieso leicht.“

I: (lachen) „Ja. Waren die für dich dann passend als du die ausgewählt hast?“

B: „Zu leicht gewesen.“

I: „Zu leicht? Also nächstes Mal dann mittel oder schwer anwählen?“

B: „Ja, mittel wär besser.“ (Proband/Probandin Z3)

Deutlich wird durch einige Aussagen, dass den Probanden und Probandinnen das Schreiben schwerer fiel als das Lesen.

I: „[...] Und fandst du die Aufgaben eher schwer oder her leicht?“

B: „Hmm, nö, find` ich eher leicht.“

I: „[...] War es eher unterschiedlich oder fandst du beides leicht, die beiden Sachen, die du gemacht hast?“

B: „Unterschiedlich.“

I: „Was fiel dir schwerer oder leichter?“

B: „Die Wörter, die man ähm schreiben musste also da eingeben musste, die waren bisschen schwerer wie lesen jetzt.“ (Proband/Probandin AJ3)

Insbesondere wird erwähnt, dass das Eingeben von Wörtern in Lücken als schwierig empfunden wurde.

Für manche Personen bestand eine große Herausforderung im Umgang mit dem Computer. Eine Person erwähnt die Herausforderung hinsichtlich der Tastatur:

„Weil die ganzen neuen Tasten. Ich mach das ja zum ersten Mal, ich kenn das ja noch nicht.“ (Proband/Probandin A3)

Eine weitere Person berichtet über Schwierigkeiten im Umgang mit der Maus:

„Ja mit der Maus musste ich mich erst anfreunden, nā? Der Pfeil wollte nicht immer dahin wo ich hin wollte.“ (Proband/Probandin AB3)

Eine Person berichtet, dass sie zum ersten mit einem Computer gearbeitet hat:

I: „Okay. Und dann, wie war zum Beispiel der Anfang oder der Einstieg für sie? Als es los ging mit dem Programm?“

B: „Schwierig, weil ich gar keinen Computer beherrsche. Ich habe das erste Mal am Computer gesessen. [...] Aber als ich das dann nachher gesagt gekriegt hatte oder gezeigt gekriegt hatte und ich wusste, wie ich was drücken musste, war es für mich kein Problem.“ (Proband/Probandin AD3)

Trotz dieser Herausforderungen hat keine Person den Test bzw. die Studie abgebrochen. Allerdings wird durch die Aussage von Proband/Probandin AD3 auch deutlich, dass die Unterstützung durch Betreuer und Betreuerinnen notwendig für die Eingewöhnung war (s. auch Erlernbarkeit).

Fazit zur Aufgabenangemessenheit

Da otu.lea sowohl im institutionellen Kontext als auch anonym genutzt werden soll, ist auf die Aufrechterhaltung der Motivation zu achten. Wären die Aufgaben nicht angemessen, hätte das vermutlich Frustration oder eine demotivierende Wirkung zur Folge. Die Möglichkeiten zur Niveau- und Dimensionswahl unterstützt die Gewährleistung der Aufgabenangemessenheit, da die Aufgaben individuell dem Kompetenzniveau entsprechend gewählt werden können. Die meisten empfanden die Aufgaben als mittelmäßig anspruchsvoll und es kann daher davon ausgegangen werden, dass die Selbsteinschätzung erfolgreich absolviert wurde. Den Aussagen der Probanden und Probandinnen zufolge kann die Aufgabenangemessenheit in otu.lea gegeben beurteilt werden, da ein Großteil der Probanden und Probandinnen die Arbeit mit otu.lea als nicht schwierig empfand.

Erwartungskonformität

otu.lea soll insbesondere für Personen mit niedriger Computererfahrung nutzbar sein. Fraglich ist, ob die Gestaltung von otu.lea für den Großteil angemessen ist. Hierfür werden die Interviews auf Aussagen hin überprüft, in denen sich Probanden und Probandinnen über die Nutzbarkeit (Nennung von Schwierigkeiten und Problemen, die sich auf die Nutzung des Programms, nicht die inhaltliche Bearbeitung der Aufgaben beziehen) 23 Personen berichten, dass sie sich gut zurechtgefunden haben und auch den Einstieg nicht als schwierig empfanden.

„Ach, eigentlich äh war das eigentlich ganz verständlich. Und umgänglich.“ (Proband/Probandin AB3)

Drei Personen geben an, sich mittelmäßig zurechtgefunden zu haben. Sieben Personen geben an, dass die Nutzung von otu.lea im Verlauf zunehmend leichter wurde, wie z. B. folgende Person (siehe auch *Erlernbarkeit*):

„Schwierig, weil ich gar keinen Computer beherrsche. Ich habe das erste Mal am Computer gesessen. [...] Aber als ich das dann nachher gesagt gekriegt hatte oder gezeigt gekriegt hatte und ich wusste wie ich was drücken musste, war es für mich kein Problem.“ (Proband/Probandin AD3)

Die Person mit dem Code AD3 berichtet, dass sie den Umgang mit dem Computer nicht beherrscht. Sie gibt zwar an, gerne mit dem Computer zu arbeiten, jedoch besitzt sie weder einen Computer noch nutzt sie einen Computer. Für eine Person sind die Wiederholungen von Instruktionen überflüssig:

„[...] bei jeder Aufgaben den Anfang, da hat sich das meiste immer meistens wiederholt. [...] Das ist dann doch bisschen einschläfernd.“ (Proband/Probandin AS3)

Zur Erläuterung ist anzumerken, dass die Teilaufgaben einer Aufgabe auf mehreren Interfaces hintereinander dargestellt werden, auf jeder Seite müssten aber dennoch die Instruktionen wieder angehört werden, auch wenn sie mit der Instruktion der vorherigen (Teil-) Aufgabe identisch sind. Die Vorannahme, die Instruktionen könnten vergessen werden, trifft auf diese Person vermutlich nicht zu. Die Instruktionen scheinen somit ausreichend oft und deutlich vorzuliegen, worauf auch die folgende Aussage einen Hinweis gibt:

I: Wünschst du dir mehr Informationen, wie das Programm zu nutzen ist?

B: Nein. Sonst wird es überladen. (Proband/Probandin AS3)

Auch die Probanden und Probandinnen mit wenig Computererfahrung fordern nicht, weitere Informationen zu erhalten. Die Person mit dem Code U3 hat zum ersten Mal mit dem Computer gearbeitet und macht folgende Aussagen:

„Für mich war, am Anfang, bin ich ganz ehrlich, als Anfänger, als jemand der noch nie da dran war, war es erst ein bisschen schwer.“ (Proband/Probandin U3)

Zudem benötigt diese Person weitere Erläuterungen zur Bedienung:

I: „Also wussten sie sofort, dass sie auf den Lautsprecher klicken mussten? Und was sie machen mussten?“

B: „Nein, mir ist das gesagt worden.“ (Proband/Probandin U3)

Diese Person weiß vorerst nicht, was sie hätte tun sollen, um mit otu.lea zu arbeiten. Aus dieser Aussage geht allerdings nicht hervor, in wie weit diese Person, die zum ersten Mal mit dem Computer gearbeitet hat, alleine mit otu.lea zurechtgekommen wäre, da sie Unterstützung erhalten hat.

Fazit zur Erwartungskonformität

Die Teilnehmenden mit wenig Computererfahrung berichten somit zwar vereinzelt von Schwierigkeiten bzw. Herausforderungen, insbesondere zu Beginn der Nutzung, jedoch wird von keiner Person ausgesagt, sie hätte otu.lea nicht nutzen können oder als zu schwierig empfunden. Auch dass eine Person (AS3) der Meinung ist, die Instruktionen müssten nicht immer wiederholt werden deutet darauf hin, dass otu.lea hinsichtlich der Instruktionen verständlich ist. Die Probanden und Probandinnen haben sich somit überwiegend erwartungskonform verhalten. Lediglich zu Beginn wurden unerwartet Funktionen nicht genutzt oder antizipiert (siehe hierzu auch *Selbstbeschreibungsfähigkeit*).

Steuerbarkeit

Hinsichtlich konkreter Charakteristika zur Selbstbeschreibung, wie sie im Kap. 8.1.2.1 zur Operationalisierung formuliert wurden, haben sich die Teilnehmenden nicht geäußert. So werden keine Aussagen darüber getroffen, selbst über Beginn oder Ende entscheiden zu wollen. Auch merkt keine Person an, dass ihr die Möglichkeit zur zwischenzeitlichen Unterbrechung fehle.

Schwierigkeiten nennen zwei Person hinsichtlich der rigiden Führung des Programms:

*„ [...] und dann bin ich auch einmal zu schnell schon weiter gegangen und dann hat er mich ja komplett rausgeschmissen und deswegen habe ich ja dann die erste Aufgabe, da hätte ich warten müssen, dass war mein Fehler. Ich war noch nicht fertig und denke da steht weiter, also gehe ich auf weiter.“
(Proband/Probandin O3)*

Hier vermutet die Person, selbst für den Absturz des Programms verantwortlich zu sein (was nicht den Tatsachen entspricht). Sie hat den Eindruck, sich zu wenig an das Programm angepasst zu haben, was wiederum ein Indiz dafür ist, dass das Programm aus der Perspektive des Nutzers/der Nutzerin eine rigide Führung aufweist.

Eine weitere Person beschreibt, dass sie gerne zu einer vorherigen Aufgabe zurückgekehrt wäre:

I: „Welche Erfahrungen haben Sie mit der Hilfefunktion gemacht?“

B: „Ja da klickt man an praktisch dann liest man alles durch und hier Fragezeichen wollt ich einmal zurück gehn hats nicht geklappt“

I: „Aber Sie haben mal draufgeklickt?“

*B: „Ja einmal. Ich wollt mal versuchen. Ich wollt nochmal zurückgehen weil ich das andere vergessen hab zu schreiben und dass hat dann nicht geklappt, das war dann schade. Das wäre dann gut. Dann hätte aber Zurück stehen müssen, wo anders. Das war dann nicht so gut. Dann bin ich weiter gegangen.“
(Proband/Probandin G3)*

Erläuternd ist hier nochmal angemerkt, dass keine Möglichkeit besteht, zwischen den Aufgaben hin- und herzuspringen, um einen Lerneffekt zu vermeiden. Das birgt gleichzeitig das Risiko, dass Aktivierungen von Schaltflächen – z. B. das Klicken auf den „Weiter“-Button – aus Versehen und nicht rückgängig gemacht werden können.

Keine Transparenz scheint für die Probanden und Probandinnen dahingehend zu bestehen, warum das Programm in seinen Funktionen begrenzt ist. Das führt zu Irritationen und womöglich auch zu Frustration.

Eine Person würde gerne selbst über die Anzahl von Wiederholungen bestimmen:

B: „Also das Programm, als ich mir das angeschaut habe...war's ok, ich würde das Programm zwar empfehlen, aber was mich auch ein bisschen natürlich auch genervt hat...das jedes Mal das Programm es... ok für manche ist es gut, für manche ist es auch nicht... aber ähm das soll sich nicht jedes Mal sich wiederholen mit ähm drücken sie auf die Taste ähm so rechts links und so...dass man auch alleine ausprobieren könnte, was es (unverständlich) sonst fand ich das Programm... fand ich das Programm gut und ich werde es auch weiterempfehlen, dass es also zu diesem ähm zu diesem Kurs auch ähm eingebracht wird, weil das ist richtig gut und ähm kann man gut lernen damit.“ (Proband/Probandin T3)

Insgesamt äußern nur wenige Personen Kritik hinsichtlich der Steuerbarkeit.

Ein weiterer Hinweis für die Steuerbarkeit ist, ob die Teilnehmenden das Programm als zu schnell bzw. zu langsam empfinden. Der Mittelwert für das Gegensatzpaar „schnell-langsam“ aus dem UEQ beträgt 0,7 (SD=1,8). Das bedeutet, die Befragten bewerteten otu.lea in der Tendenz als langsam.

Dass keine Aussagen über die Möglichkeit, über Beginn oder Ende entscheiden zu können, getroffen werden, kann in der für die Teilnehmenden ver-

mutlich strikt erscheinenden Versuchsanordnung begründet sein: Möglicherweise ziehen sie nicht die Möglichkeit in Betracht, über Beginn, Ende und Unterbrechung entscheiden zu wollen. Das Bedürfnis, darüber selbst zu bestimmen, wird vermutlich erst in der selbstständigen Nutzung und ohne kontrollierend empfundene Personen (Betreuer und Betreuerinnen) auftreten.

Die Tendenz, otu.lea als langsam zu bewerten ist minimal und es kann behauptet werden, dass otu.lea als weder zu schnell noch zu langsam beurteilt wird¹⁰⁶.

Fazit zur Steuerbarkeit

Bezüglich des Re-Designs zu überlegen, ob näher erläutert wird, warum es nicht möglich ist, zwischen den Aufgaben hin- und herzuspringen. Die Ergebnisse geben nur bedingt Aufschluss über die Steuerbarkeit. Weitere Evaluationen, beispielsweise in Form von Interviews mit gezielteren Fragen zur Steuerbarkeit, wären für eine stärkere Aussagekraft sinnvoll. Auf der Grundlage der hier dargestellten Ergebnisse ist allerdings davon auszugehen, dass die Steuerbarkeit bei otu.lea vorhanden ist.

Perspektivübernahme

Die Aussagen zur Verständlichkeit der Instruktionen sind recht homogen, indem die meisten Probanden und Probandinnen diese als einfach und verständlich bezeichnen, exemplarisch für den Großteil der Probanden und Probandinnen sei hier eine Aussage aufgeführt:

„Und bei den Texten war eigentlich klipp und klar, alles klar geschrieben. Alles konnte ich super verstehen.“ (Proband/Probandin H3)

Allerdings wird auch von drei Personen angemerkt, dass die Texte teilweise als zu lang empfunden werden, was z. B. folgende Aussage verdeutlicht:

„Weil manchmal, wenn du dann so lange Texte hast, dann hast du manchmal irgendwie so ‘n kleinen... Dreher, dass du’s dann nicht mehr weißt. Und deswegen fand ich das dann halt gut, dass man die dann wieder anklicken konnte.“ (Proband/Probandin J3)

¹⁰⁶ Allerdings ist dieses Ergebnis aus dem UEQ nicht mit Sicherheit als aussagekräftig zu bewerten, da das Gegensatzpaar „schnell-langsam“ der Skala „Effizienz“ angehört, die einen unbefriedigenden Cronbachs α aufweist. Das Ergebnis wird aber dennoch einbezogen, da eine Missinterpretation des Gegensatzpaares „schnell-langsam“ als unwahrscheinlich angesehen wird.

Problematisch ist für eine Person die Schriftgröße:

„Da hatte ich sehr viel Schwierigkeiten das überhaupt zu lesen, dass schwamm alles in einander, das war nicht gut. [...] Aber es war auch nur eine Seite. [...] Die hab ich gleich weiter geklickt, da hab ich nicht dran gearbeitet.“ (Proband/Probandin AD3)

Die Schriftgröße eines Stimulus ist für einen Proband/eine Probandin zu klein, um den Text lesen und die Aufgabe anschließend bearbeiten zu können. Unverständlich ist für eine Person die Instruktion hinsichtlich der Eingabe von Satzzeichen:

B: „Und deshalb habe ich mir da das erstmal was überlegt, aber das erste Mal wollte er ja nicht den Punkt haben. Und dann hab ich das mit dem Nächsten versucht. In dem Schreiben stand nämlich drinne, das das die Grauen sind. [...] Und das hat sie am Dings nicht erzählt. Sie hat da nur was erzählt vom quadratischen Kästchen.“

I: „Würden Sie denn das wünschen, dass das verbessert wird?“

B: „Unbedingt [...] Weil, du kannst ja nicht nen Text lesen, es gibt ja welche, die können nen Text lesen und hören. [...] Und wenn da im Text drin steht, die sind grau [...] und mit den Ohren hörst du [...] quadratisch. Dann muss da auch die Farbe hin.“ (Proband/Probandin AN3)

Hier wird deutlich, dass die Instruktion nicht eindeutig mit der Aufgabendarstellung (der Farbe der „Kästchen“) übereinstimmt.

In den meisten Aussagen zur Verständlichkeit von otu.lea wird die Testumgebung als leicht zu verstehen bezeichnet. Vereinzelt merken Probanden und Probandinnen Schwierigkeiten und Herausforderungen an. Die Texte werden vereinzelt als zu lang bewertet, die Schrift empfindet eine Person als zu klein. Eine Transferleistung zur Überbrückung der mangelnden Übereinstimmung zwischen Instruktion und Aufgabe kann von Probanden und Probandinnen nicht erwartet werden.

Fazit zur Perspektivübernahme

Insgesamt ist der Instruktionstext für die Aufgaben, in denen Satzzeichen gesetzt werden müssen und „Kästchen“ erwähnt werden, zu überarbeiten. Ein Re-

Design könnte zudem die Einbettung einer Möglichkeit zur Schriftgrößenveränderung beinhalten. Die Perspektivübernahme ist somit bedingt gegeben.

Prozessangemessenheit

Für die Beurteilung der Prozessangemessenheit werden Interviewaussagen aus dem Experten-/Expertinnengespräch sowie Aussagen der Probanden und Probandinnen herangezogen.

In dem Gespräch mit Kursleitenden wird besprochen, in wie weit der Einsatz von otu.lea für die Alphabetisierungskräfte in ihren Kursen vorstellbar ist und ob sie darin eine angemessene Unterstützung ihrer Tätigkeit sehen. Im Gespräch wird mehrfach erwähnt, dass eine individuelle Betreuung der Kursteilnehmenden notwendig ist.

A: „Und wir müssen immer drauf eingehen, wie die jetzt drauf sind und was sie brauchen. Und es gibt ja auch keine festen, uhm, Vorgaben, die wir da haben, sondern wir müssen auf jeden Einzelnen eingehen und ihn da abholen, wo er gerade ist.“

Von dem Kursleiter/der Kursleiterin wird die Herausforderung beschrieben, dass auf jeden Kursteilnehmer/jede Kursteilnehmerin entsprechend ihrer/seiner Kompetenzen eingegangen werden muss. Hier wird bereits eine potenzielle Passung zwischen einer Herausforderung in der Alphabetisierung und otu.lea deutlich: Die otu.lea-Diagnostik ermöglicht durch die differenzierte und individuelle Auswertung die Anknüpfung an die individuellen Kompetenzen.

In der folgenden Aussage des gleichen Kursleiters/der gleichen Kursleiterin wird zudem auf die individuelle Betreuung in der Kursarbeit hingewiesen. Die/der Kursleitende beschreibt explizit eine potenzielle Unterstützungsmöglichkeit durch otu.lea.

A: „Wenn man sie einzeln hinsetzen kann, und dann kann man ja hinterher sich mit den Leuten beschäftigen. Wenn einer beschäftigt ist und dann kannst du mit dem anderen was machen. [...] aber wenn sie sich selbst damit beschäftigen und auch nicht immer gleich Hilfe kriegen. Das finde ich gut.“

Einerseits sieht der/die Kursleitende eine potenzielle Entlastung, da er/sie eine Person individuell betreuen und ein anderer Kursteilnehmer/eine andere Kursteilnehmerin parallel mit otu.lea arbeiten kann. Somit muss die/der Kursleitende nicht zwei Personen gleichzeitig betreuen und eine Person individuell unterstüt-

zen. Andererseits sieht der Kursleiter/die Kursleiterin die Teilnehmenden in ihrer Selbstständigkeit gefordert.

Eine weitere Passung zwischen otu.lea und einer Herausforderung in der Alphabetisierung wird in folgender Passage deutlich.

B: „[...] andere Herausforderung ist, dass es dadurch ja auch, gibt ja wenig bisher wenig erwachsenengerechtes Unterrichtsmaterial.“

A: „Ja, das fehlt uns also wirklich.“

[...]

A: „Und wir (...) bei Kindern und das passt dann vom Thema nicht. Also wenn wir da unsere erwachsenen Männer da haben und dann kommt da auf einmal Puppe.“

[...]

A: „Das geht gar nicht.“

Zwei Kursleitende erwähnen, dass bisher wenig erwachsenengerechte Materialien vorhanden sind. Die Nutzung von Material, welches für Kinder entwickelt wurde, sei unpassend („Das geht gar nicht.“). Daraufhin wird beschrieben, wie otu.lea dem Anspruch an erwachsenengerechtem Material begegnet.

A: „Also der Text ist ja ganz gut oder die Geschichten sind ja ganz gut, weil sie so aus der Arbeitswelt angegliedert sind und eher (Hilfsberufe), würde ich mal sagen, falls es noch heute so was gibt, heutzutage. Also von daher ist es, sind die Geschichten ganz gut, weil sie aus so einer Erwachsenenarbeitswelt ist, die eine, die alleinerziehend ist und Kind hat und so. Das fand ich ganz gut.“

Konkret berichtet beispielsweise eine Kursleiterin/ein Kursleiter zudem, dass eine Kursteilnehmerin/ein Kursteilnehmer als Reinigungskraft arbeitet (Zeile 226). Somit scheint otu.lea auch aus der Perspektive der Kursleitenden die Arbeits- und Lebenswelt der Kursteilnehmenden angemessen widerzuspiegeln. In einer weiteren Aussage wird die Wichtigkeit einer klaren Struktur betont.

D: „Und es ist ganz wichtig, dass es klar gegliedert ist. Dass sie selbst entscheiden können, was ist jetzt als nächstes dran oder so und, dass sie es sofort erkennen, neh.“

Auch wird hier erwähnt, dass die Möglichkeit der selbstständigen Entscheidung für die Teilnehmenden wichtig ist.

Ob die Teilnehmenden der Meinung der Kursleitenden nach *otu.lea* auch selbstständig außerhalb der Alphabetisierungskurse nutzen würden, wird von einer Person positiv beantwortet.

D: „Also zu dieser ich-will-lernen.de, da hab ich, die zwei fortgeschrittenen Teilnehmer, junge Männer, die auch Arbeit haben, den habe ich das gesagt und den anderen dann auch. Und ich denke, dass sie da auch manchmal reingucken. Also so ganz verloren geht es nicht.“

Das Lernportal *ich-will-lernen.de* hat in der Alphabetisierungsbranche einen hohen Bekanntheitsgrad. Die Kursleiterin/Der Kursleiter geht davon aus, dass Teilnehmende das Portal auch selbstständig außerhalb des Alphabetisierungskurses nutzen und folgert daraus, dass dies auch für *otu.lea* gelten könnte.

Ein Kursteilnehmer/Eine Kursteilnehmerin merkt an, dass eine Verknüpfung von *otu.lea* und dem Lernportal sinnvoll wäre.

I: „Ähm, würdest du gerne öfter mit dem Programm arbeiten?“

B: „Eigentlich ja, das kann man eigentlich, wenn man halt die (anderen Sachen?) einfügt. Alles in einem (packt?), dass man sich da zwischen entscheiden kann, Rechnen oder wie auch immer. Weil ähh, das Programm hier, das „Ich will Lernen“, da arbeite ich ja auch dran zwischenzeitlich. Und deswegen wäre das auch ganz gut, wenn man da auch Rechnen mit einfügen könnte.“ (Proband/Probandin R3)

Die Person sieht einen potenziellen Mehrwert für die Alphabetisierungsarbeit in der Verzahnung beider Programme durch die Abdeckung der gleichen Kompetenzbereiche und die Integration in die Alphabetisierungspraxis.

Fazit zur Prozessangemessenheit

Die Aussagen deuten mehrfach auf einen angemessenen Einsatz von otu.lea sowohl in Alphabetisierungskursen als auch außerhalb der Kurse hin:

- Anknüpfung an die individuellen Kompetenzen
- durch die selbstständige Nutzungsmöglichkeit entsteht die Möglichkeit sowohl einer individuellen Betreuung als auch der Förderung der Selbstständigkeit der Teilnehmenden
- erwachsenengerechte Gestaltung der Online-Testumgebung
- klare Struktur

Die Prozessangemessenheit ist somit als erfüllt zu beurteilen.

Joy of Use

Ein Großteil der Probanden und Probandinnen kann sich vorstellen, mit dem Programm wieder zu arbeiten. So bestätigt zum Beispiel eine Person, dass sie sich auch vorstellen könnte, das Programm zu Hause zu nutzen:

I: „Okay. Und würdest du gerne öfters mit diesem Programm arbeiten?“

B: „Joa, so.“

I: „Auch zuhause?“

B: „Joa.“ (Proband/Probandin E3)

Eine Person bewertet otu.lea als spannend:

B: „Ich fand das... ja sehr spannend.“ (Proband/Probandin D3)

Auch das Storyboard trägt zur Motivation bei:

B: „Ja, in einen ja. Zum Beispiel bei der Reinigungskraft, zum Beispiel hat die mich da sehr weil... ich war auch eine Reinigungskraft und ?hier bin ich? Daher hat ja das auch ganz viel Spaß dann gemacht. Das dann überhaupt die Aufgaben zu lösen.“ (Proband/Probandin AV3)

Eine Person (47 Jahre) ist allerdings der Meinung, dass otu.lea eher etwas für jüngere Personen ist:

B: „Ja, wie soll ich es sagen? Ich würd es vielleicht nicht benutzen, so sage ich mal, wenn man ehrlich ist.“ (lacht)

I: „Warum?“

*B: „Weiß ich nicht. Es würd mir so nicht liegen, sag ich mal. Es ist mehr für jüngere Leute, sag ich mal. Es ist jetzt, ob das blöd ist, weiß ich nicht ja aber.“
(Proband/Probandin AQ3)*

Bei 13 Personen kommt es im Verlauf der Nutzung von otu.lea zu einer Motivationssteigerung. Davon berichten sechs Personen, dass sie anfangs nicht motiviert waren, otu.lea auszuprobieren; sieben Personen sagen aus, zu Beginn nervös oder ängstlich gewesen zu sein:

I: „[...] Und, genau jetzt frag ich dich noch wie motiviert warst du das Programm zu nutzen?“

B: „Geht.“

I: „Warum?“

B: „Ja [...] Angst was auf sich zukommt und so.“

I: „Ok. Und jetzt im Nachhinein?“

B: „Besser.“ (Proband/Probandin AF3)

Ein Teilnehmer/Eine Teilnehmerin beschreibt die Herausforderung, die in der erstmaligen Nutzung eines Programmes liegt:

I: „Ja und wie motiviert warst du das Programm zu nutzen?“

B: „Hmm, nicht so.“

I: „Warum?“

B: „Weil das... weil man das erstmal nicht kennt, aber wenn man erstmal reinkommt dann eigentlich doch.“

I: „Aha. Wieso dann doch?“

B: „Irgendwann macht das dann doch Spaß, wenn man sich das durchliest und das machen kann, das lernen kann.“ (Proband/Probandin E3)

Die Nutzung des unbekannten Programms löst bei der Person E3 vorerst Unbehagen aus. Allerdings ist dies nicht auf eine geringe Computererfahrung zurückzuführen, denn sie besitzt einen Computer und nutzt diesen auch regelmäßig. Die Ergebnisse des UEQ zeigen, dass die Attraktivität von otu.lea im Mittel mit 2,77 und somit als sehr ausgeprägt bewertet wird. Die Attraktivität setzt sich zusammen aus den Gegensatzpaaren „unerfreulich – erfreulich“, „gut –

schlecht“, „abstoßend – anziehend“, „schön – hässlich“ sowie „sympathisch – unsympathisch“.

Die Aussage, dass otu.lea eher für jüngere Personen geeignet ist, ist vermutlich darin begründet, dass diese Person per se die Nutzung von Computern tendenziell jüngeren Personen zuschreibt. Möglicherweise hat diese Person die Meinung über sich, nicht gut mit dem Computer arbeiten zu können. Die Angaben zur Computernutzung zeigen, dass sie eher wenig den Computer und das Internet nutzt.

Dass Personen anfangs nicht so motiviert sind, otu.lea zu nutzen, kann darin begründet sein, dass sie Computern gegenüber vorerst skeptisch sind oder auch neuen Dingen, mit denen sie nicht vertraut sind, mit Vorsicht begegnen. Auch Personen mit viel Computererfahrung äußern mangelnde Motivation aufgrund dessen, dass sie das Programm nicht kennen. Unter Berücksichtigung dieser Aussagen ist daher insbesondere auf einen leichten Einstieg zu achten, der den Abbau von Hemmschwellen und die Steigerung der Motivation ermöglicht.

Fazit zu Joy of Use

Zu Beginn ist auf einen leichten Einstieg zu achten, der zudem eine möglichst breite Nutzer- und Nutzerinnenschaft anspricht. Einerseits können dadurch Hemmschwellen abgebaut und andererseits die Motivation aufrecht erhalten bzw. gesteigert werden.

Zusammenfassend ist nach den eben dargestellten Interviewaussagen und den Ergebnissen des UEQs bezüglich der Nutzung von otu.lea festzuhalten, dass die Probanden und Probandinnen zufrieden zu sein scheinen und gerne mit otu.lea arbeiten.

Die Heuristik Joy of Use ist somit als erfüllt zu beurteilen.

Interkulturelle Aspekte

Die Meinungen zum Storyboard sind heterogen. Das Storyboard besteht aus Comic-Zeichnungen, die Figuren sind – also keine realen Abbildung, sondern gezeichnet.

Einige Probanden und Probandinnen finden die Geschichten interessant und spannend:

„Hm, ja. Also da ist eben halt ja... wie gesagt, äh. Das sie eben halt... ja durch... durch die Geschichte, was da jetzt so auch noch passiert und dann möchte man dann wissen, was dann bei der nächsten Geschichte ist und ja... was man dann da noch so lösen muss und auch das Bild von dem, ach... sehr gut was die da so da hergestellt haben.“ (Proband/Probandin AV3)

Eine Person kann auch konkret benennen, was sie an den Geschichten spannend empfindet:

I: „Zu diesen Personen. Ähm wie fandst du das denn?“

B: „Ich fand das ja sehr spannend.“

I: „Fandst spannend? Was fandst denn spannend daran?“

B: „Das mit der Auszubildenden. (lacht) Das die frech war.“ (Proband/Probandin B3)

Zum Storyboard äußern vereinzelt Probanden und Probandinnen konkret, womit sie sich identifizieren können:

I: „Ja und konnten sie sich denn gut in diese Personen hineinversetzen?“

B: „Ja.“

I: „Ja?“

B: „Ja.“

I: „Inwiefern?“

B: „Beispiel bei der Reinigungskraft (lacht), weil ich bin auch Reinigungs(kraft?), ich arbeite auch in der Reinigung, ich mache auch () und Treppenhäuser, und daher kann ich mich in ihr gut reinversetzen wenn sie zu spät gekommen ist und dann hat sie noch ne Tochter zuhause und dann musste sie mit dem Fahrrad hinfahren. Doch ich kann mich in solche Sachen sehr gut reinversetzen und das ist das Spannende denn da, weil denn (will ich?) wissen wie das Ende ist.“ (Proband/Probandin A3)

Einer weiteren Person gefällt die Geschichte über den Bauarbeiter, weil sie selbst in diesem Bereich tätig ist:

I: „Mh, aber, aber das, aber warum war das mit dem Bau denn dann Ihre Lieblingsgeschichte?“

B: „Weil ich sowieso auf dem Bau bin, und dann hört man das ja so ein bisschen.“ (Proband/Probandin F3)

Die Identifikation findet dabei nicht unbedingt über eine Berufsbezeichnung, sondern über gemeinsame Tätigkeiten statt:

I: „Was fällt Ihnen zu den Geschichten und den Figuren im Programm ein? Konnten Sie sich in die Personen hineinversetzen?“

B: „Ja. (lacht)“

I: „Sehr gut. Bei der Köchin dann.“

B: „Ja, genau (lacht).“

I: „Sind Sie Köchin?“

B: „Ne, ich lern das jetzt, ja.“ (Proband/Probandin AX3)

Hingegen empfinden Probanden und Probandinnen die dargestellten Charaktere vereinzelt als eher befremdlich:

I: „[...] zu den Geschichten und den Figuren, ähm, wir hatten ja hier immer so Anfangsgeschichten und die Figuren. Was fällt dir dazu ein?“

B: „Komisch [...] Weil sie sehr komisch aussehen, als von... ich weiß nicht, Lernbüchern.“ (Proband/Probandin V3)

Die Frage, ob eher Zeichnungen oder reale Fotos bevorzugt würden, wird heterogen beantwortet. So sprechen sich zehn Personen eindeutig für die Comic-Zeichnungen aus, elf Personen bevorzugen Fotos von realen Personen. Die weiteren Personen äußern sich nicht oder sie haben keine Präferenz.

Begründet wird die Präferenz für Zeichnungen beispielsweise in dem Bedürfnis, die Privatsphäre zu wahren:

B: „Ich finde Zeichnungen gut, ja das finde ich gut.“

I: „Können Sie sagen, warum Sie die Zeichnungen besser finden, als die Fotos?“

B: „Weil das ist ja Privatsphäre.“

I: „Ah ok.“ (Proband/Proband Q3)

Als weiterer Grund für die Präferenz der Zeichnungen wird die Erkennbarkeit genannt:

I: „Genau das sind beides Reinigungskräfte aber wenn sie jetzt so die Auswahl hätten sie können sich aussuchen ob sie eher lieber Zeichnungen sehen

würden oder reale Fotos würden sie lieber reale Fotos oder eher Zeichnungen sehen.“

B: „Zeichnungen.“

I: „Lieber Zeichnungen können sie auch sagen warum?“

B: „Kann man besser erkennen.“ (Proband/Probandin AG3)

Ein Grund für die Befürwortung von Fotos ist beispielsweise die bessere Erkennbarkeit von Emotionen:

I: Ja ok. So jetzt habe ich noch eine Frage und zwar haben wir in den Programm, Figuren wie... (zeigt Screenshot). So eine Putzfrau hatten Sie vielleicht oder einen KFZ-Mechaniker. Und das haben wir ja dadrin. Und das sind ja solche Zeichnungen. Richtig?

B: Ehem (Zustimmung)

I: Jetzt haben wir jetzt mal gedacht, wir fotografieren diese Personen mal ab in Echt. Was fänden Sie da besser? Würden Sie sagen, lasst das mal, dass ist schön oder würden Sie sagen, ich möchte lieber Fotos?

B: Kann man sagen, dass besser, das wird gemalt. Kann man sehen wie Gefühle, wie Fotos

I: Fotos finden Sie besser? Weil man das besser sieht?

B: Ja (Proband/Probandin AP3)

Auch folgende Person nennt einen ähnlichen Grund, indem sie die Zeichnungen als „roboterhaft“ beurteilt:

I: Und jetzt hätten wir eine Frage noch, bevorzugen Sie eher die realen Bilder oder die Comics die wir jetzt hatten?

B: So ist besser

I: Also so, real finden Sie besser?

B: Ehem (Zustimmung). Das wie Roboter alles. (Proband/Probandin AE3)

Da mit dem Begriff „Roboter“ geläufig „emotionslos“ assoziiert wird, ist zu vermuten, dass auch diese Person Abbildungen bevorzugt, in denen Emotionen zu erkennen sind.

Vereinzelte liefern die Aussagen Hinweise darauf, dass das Identifikationspotenzial gegeben ist. Die Aussagen beziehen sich zwar auf das Storyboard, allerdings werden in keinem der Gespräche interkulturelle Aspekte wie beispielsweise die Hautfarbe und die Nationalität explizit thematisiert. Anzumerken ist allerdings, dass auch nicht explizit danach gefragt wird. Dass kein Proband/keine Probandin explizit die Interkulturalität des Storyboards kommentiert kann darin begründet sein, dass sie es entweder nicht interessiert, es ihnen nicht auffällt oder sie sich im Gespräch nicht daran erinnern.

Fazit zu interkulturellen Aspekten

Ob die interkulturellen Aspekte berücksichtigt werden, kann nicht erschöpfend beantwortet werden. Um die Heuristik „Interkulturelle Aspekte“ beurteilen zu können, müssten weitere Studien – vorzugsweise Interviews, in denen explizit nach interkulturellen Aspekten gefragt wird – durchgeführt werden.

Bearbeitung der Frage F1b: „Gibt es Hinweise auf die Notwendigkeit, neue Heuristiken zu formulieren?“

Im Folgenden werden Beobachtungen und Ergebnisse angeführt, die auf Herausforderungen im Umgang mit der Online-Diagnostik hindeuten, aber nicht durch die Anwendung der 15 Heuristiken (vgl. Kap. 2.6.5) bedient werden:

- Die Online-Diagnostik wird auch von Personen ohne Computererfahrung genutzt (Interviewaussage U3)
- Nervosität und Hemmungen können die Aufmerksamkeit und Konzentration beeinträchtigen (Analyse Zyklus 1 Kap. 6.3 und Analyse Zyklus 2 Kap. 7.3)
- Probanden/Probandinnen haben zu Beginn mehr Schwierigkeiten und benötigen zu Beginn mehr Unterstützung als am Ende (s. Kap. 8.3.3)

Für die Nutzung der Online-Diagnostik muss somit berücksichtigt werden, dass sie auch von Personen ohne jegliche Computererfahrung genutzt wird. Personen ohne Computererfahrung haben womöglich kein Vertrauen in die Testumgebung und haben Hemmschwellen, sich explorativ zu bewegen und Funktionen auszuprobieren. Nervosität kann ebenso zum Aufbau von Hemmschwellen führen. Zudem können durch die Nervosität kognitive Ressourcen gebunden werden, so dass weniger Gedächtniskapazität für die inhaltliche Bearbeitung der Aufgabe zur Verfügung steht. Des Weiteren hat die Studie des dritten Zyklus gezeigt, dass insbesondere zu Beginn Unterstützung notwendig ist. Möglicherweise ist dies

ebenfalls darin begründet, dass sich die Nutzer und Nutzerinnen nicht trauen, explorativ vorzugehen und Angst haben, etwas falsch zu machen.

Vor dem Hintergrund dieser Ergebnisse und Beobachtungen erscheint es sinnvoll, eine Übungssequenz vor der kompetenzdiagnostisch relevanten Bearbeitung der Aufgaben zu ermöglichen¹⁰⁷. Einerseits kann eine Übungssequenz dazu beitragen, Testangst und Nervosität abzubauen. Andererseits würde die Übungssequenz den insbesondere zu Beginn vorhandenen Unterstützungsbedarf aufgreifen, indem die Probanden und Probandinnen die Umgebung explorieren können, ohne die Befürchtung, dass sich ihre Aktionen auf die Performanzmessung auswirken. Im Kontext computerbasierten Testens kann daher eine weitere Usability-Heuristik lauten:

Übungssequenz: Bei der Gestaltung von computerbasierten Testumgebungen sind Übungssequenzen vorzuschalten, durch die sich Teilnehmende mit grundsätzlichen Funktionen vertraut machen können.

Im Verlauf der weiteren Ergebnisdarstellung sind diese Überlegungen ggf. wieder aufzugreifen.

8.3.4 Zusammenfassung und Diskussion der Ergebnisse zu den Usability-Heuristiken

Die evaluierten Usability-Heuristiken sind in otu.lea zu einem Großteil berücksichtigt. Folgende Modifikationsbedarfe sind aus den Studien abzuleiten. Vereinzelte ergeben sich auch Modifikationsmöglichkeiten, auch wenn die Heuristik grundsätzlich als erfüllt bewertet wird.

¹⁰⁷ Diese Überlegungen sind allerdings ausschließlich auf den Einsatz von computerbasierten Tests zu beziehen und nicht generell auf die Nutzung von Webseiten übertragbar.

Usability-Heuristik	Erfüllt (ja/nein/teilweise)	Modifikationsbedarf und Bemerkungen
Erlernbarkeit	ja	Einbindung von Pop Up-Fenstern nach einem längeren Zeitraum (ca. 5 Minuten) ohne Interaktion
Selbstbeschreibungsfähigkeit	teilweise	Deutlicherer Hinweis auf die Hilfefunktion, Beenden der Hilfefunktion einfacher darstellen; Wechsel zwischen Teilaufgaben transparenter und deutlicher darstellen; selbsterklärende Gestaltung der Login-Sequenz durch Erläuterungen zur Code-Eingabe und iterativen Seitenaufbau
Wahrnehmungssteuerung	nicht beurteilbar	Durchführung weiterer Evaluationen (beispielsweise ein Eye Tracking-Experiment)
Aufgabengemessenheit	ja	keine
Erwartungskonformität	Ja	keine
Steuerbarkeit	Ja	Erläuterung für die Teilnehmenden, dass zwischen den Aufgaben nicht hin- und hergesprungen werden kann
Perspektivübernahme	teilweise	Überarbeitung von Instruktionstexten, Implementation der Möglichkeit zur Veränderung der Schriftgröße
Prozessangemessenheit	ja	keine
Joy of Use	ja	keine
Interkulturelle Aspekte	nicht beurteilbar	Durchführung weiterer Evaluationsstudien, beispielsweise mit Interviews

Tabelle 36: Erfüllung der Heuristiken und Modifikationsbedarfe für otu.lea

Priorität liegt auf den Modifikationsbedarfen zur Selbstbeschreibungsfähigkeit. Insbesondere der Login-Bereich ist zu überarbeiten, da die Bearbeitung des Login-Bereichs erst die Teilnahme an otu.lea ermöglicht. Auch ist gerade zu Beginn darauf zu achten, dass nicht aufgrund mangelnder Verständlichkeit die Teilnahmemotivation sinkt und potenzielle Teilnehmende den Test abbrechen bevor sie mit der Aufgabenbearbeitung begonnen haben. Da die Ergebnisse stark kontextabhängig sind, ist eine grundsätzliche Aussage, in wie weit die überprüften Heuristiken für funktionale Analphabeten und Analphabetinnen relevant und/oder ggf. zu modifizieren sind, nur bedingt möglich. Auch sind die Ergebnisse vor dem Hintergrund der Stichprobenauswahl zu betrachten: Dadurch, dass ausschließlich Personen aus Alphabetisierungskursen an den Studien teilgenommen haben, ist von einem positiven Selektionseffekt auszugehen: Die Teilnehmenden nehmen freiwillig an den Alphabetisierungskursen teil, haben zum

Zeitpunkt der Studie bereits Lerneinheiten zur Literalitätsförderung durchlaufen und sind höchstwahrscheinlich lernmotiviert. Daher sind die Ergebnisse nicht auf die Grundgesamtheit funktionaler Analphabeten und Analphabetinnen zu übertragen. Womöglich haben die teilgenommenen Personen im Durchschnitt bessere Performanzwerte erzielt als die Grundgesamtheit¹⁰⁸. Allerdings zeigen die Ergebnisse zur Computererfahrung und -kompetenz ein heterogenes Bild, so dass dahingehend kein positiver Selektionseffekt zu vermuten ist.

8.3.5 *Ergebnisdarstellung zur Computererfahrung und -kompetenz*

Im Folgenden werden die Ergebnisse den Fragen F2a und F2b zugeordnet dargestellt.

Bearbeitung der Frage F2a: „Haben die Computererfahrung und -kompetenz einen Einfluss auf die Bedienbarkeit?“

Zur Beantwortung der Frage 2a und der Überprüfung der Hypothesen H_{0a1} und H_{0a2} sind die Ergebnisse zur Computererfahrung¹⁰⁹, die eingeforderte Unterstützung sowie die Aufgabenbearbeitungszeit zu berücksichtigen¹¹⁰. Hierfür werden t-Tests mit unabhängigen Stichproben berechnet¹¹¹.

Die Ergebnisse zur Überprüfung eines Effekts der Computererfahrung auf die Unterstützung zeigen signifikante Werte¹¹²: Es bestehen sowohl zwischen der Computererfahrung und der eingeforderten Unterstützung ($p=.02$) als auch zwischen der Computererfahrung und der Unterstützung insgesamt (eingefordert und unaufgefordert) ($p=.00$) signifikante Effekte.

¹⁰⁸ Die in der leo.-Studie generierten Ergebnisse zum funktionalen Analphabetismus wurden mit anderen Aufgaben als in otu.lea erhoben. Zudem hatte die leo.-Studie die Zuordnung zu den Alpha-Level zum Ziel, was in der leo.-Diagnostik nicht vorgesehen und eine Angabe in diesem Kontext dazu somit nicht möglich ist.

¹⁰⁹ Die Computererfahrung wird dabei als kategoriale Variable verwendet, indem die Probanden und Probandinnen den Kategorien viel/wenig Computererfahrung zugeordnet werden.

¹¹⁰ Die Ergebnisse zur Computerkompetenz werden aufgrund der zu geringen Varianz nicht einbezogen.

¹¹¹ Die Voraussetzung für den t-Test wurden überprüft. Die Bedingung der Normalverteilung ist für die Computererfahrung erfüllt. Die Werte für die eingeforderte Unterstützung weichen von der Normalverteilung ab: Die Nennungen für keine Unterstützung treten bei mehr als der Hälfte der Probanden und Probandinnen auf. Da es sich beim t-Test aber um einen robusten Test handelt, können die Abweichungen toleriert werden (Büning, 1991).

¹¹² Die Power-Analyse ergibt bei einer angenommenen mittleren bis großen Effektstärke von 0,7 (vgl. z. B. Bortz & Döring, 2006, S. 642) und einem angenommenen Alpha-Fehler von 0,1 einen Wert von 0,8. Somit wird das Signifikanzniveau auf 0,1 festgelegt.

Die folgende Tabelle zeigt die Anzahl der eingeforderten Unterstützung:

		Anzahl der eingeforderten Unterstützung										Gesamtsumme
		1	2	3	4	6	8	15	19	20	22	
Computererfahrung	wenig	5	3	4	1	1	2	1	1	1	1	20
	viel	5	0	3	1	0	0	0	0	0	0	9
Gesamtsumme		10	3	7	2	1	2	1	1	1	1	29

Tabelle 37: Computererfahrung und Anzahl der eingeforderten Unterstützung

Insgesamt fordern Personen mit wenig Computererfahrung 20-mal Unterstützung ein, wohingegen Personen mit viel Computererfahrung nur neunmal Unterstützung einfordern. Auch hinsichtlich der Unterstützung insgesamt erhalten Personen mit wenig Computererfahrung öfter Unterstützung als Personen mit viel Computererfahrung:

		Anzahl der Unterstützung gesamt																		Gesamtsumme
		0	1	2	3	4	5	6	9	11	13	14	15	29	30	32	36	47		
Computererfahrung	wenig	3	6	2	3	2	0	1	1	1	1	1	1	1	1	1	1	1	27	
	viel	8	3	6	1	1	2	1	0	0	0	0	0	0	0	0	0	0	22	
Gesamtsumme		11	9	8	4	3	2	2	1	1	1	1	1	1	1	1	1	1	49	

Tabelle 38: Computererfahrung und Unterstützung gesamt

Personen mit wenig Computererfahrung erhalten jeweils mehr Unterstützung im Vergleich zu Personen mit viel Computererfahrung.

Somit ist die H_{0a1} zugunsten der H_{1a1} zu verwerfen und es gilt:

Es besteht ein Zusammenhang zwischen der Computererfahrung und der eingeforderten Unterstützung: Personen mit wenig Computererfahrung werden öfter unterstützt als Personen mit viel Computererfahrung.

Zur Überprüfung eines möglichen Zusammenhangs zwischen der Computererfahrung und der Aufgabenbearbeitungszeit deuten die Mittelwerte für die Aufgabenbearbeitungszeit (differenziert nach den Kategorien viel und wenig Computererfahrung) auf einen Zusammenhang hin: Für 29 von 31 Aufgaben ist die durchschnittliche Aufgabenbearbeitungszeit für Personen mit wenig Computererfahrung höher. Lediglich für die Aufgaben 1.5.2 und 1.6.1 ist dies nicht der Fall. Für diese Aufgaben sei angemerkt, dass sie lediglich von zwei Personen mit

wenig und einer Person mit viel Computererfahrung bearbeitet wurden. Dieses Verhältnis der durchschnittlichen Aufgabebearbeitungszeit zwischen Personen mit wenig und mit viel Computererfahrung kann somit aufgrund der niedrigen Fallzahl als nicht aussagekräftig angesehen werden.

Für die Überprüfung des Zusammenhangs zwischen der Computererfahrung und der Aufgabebearbeitungszeit ergibt der t-Test für unabhängige Stichproben signifikante Ergebnisse für 9 von 31 Aufgaben¹¹³: 1.2.1 ($p=.02$), 2.1.02_II ($p=.04$), 2.2.01_III ($p=.09$), 2.2.02_II ($p=.05$), 2.2.02_III ($p=.03$), 2.3.01_I ($p=.01$), 2.3.03_III ($p=.01$), 2.3.03_IV ($p=.08$) und 2.4.01_I ($p=.05$).

Somit gilt für die genannten 9 von 31 Aufgaben: Personen mit wenig Computererfahrung benötigen mehr Bearbeitungszeit als Personen mit viel Computererfahrung. Anzumerken ist, dass im Vergleich zur Dimension Lesen mit ausschließlich Multiple Choice-Aufgaben wesentlich mehr signifikante Werte im Bereich Schreiben mit halboffenen und offenen Antwortformaten vorhanden sind. Um zu überprüfen, ob eine mögliche Ursache in bestimmten Formaten zu finden ist, werden die formatbezogenen durchschnittlichen Aufgabebearbeitungszeiten in der Berechnung berücksichtigt.¹¹⁴

Unter Berücksichtigung des Aufgabenformats bestehen Effekte zwischen der Computererfahrung und der durchschnittlichen Aufgabebearbeitungszeit bei 4 von 8 dimensionsbezogenen Formaten: *offenes Format Schreiben Einfach* ($p=.09$), *halboffenes Format Schreiben Mittel* ($p=.01$), *offenes Format Schreiben mittel* ($p=.02$) und *Korrekturformat Schreiben mittel* ($p=.07$).

Somit tritt bei den „einfachen“ Formaten (vgl. auch Kap. 2.2.6 zu Itemformaten) und auf niedrigem Niveau sowohl in der Dimension Lesen (Single Choice Lesen einfach) als auch in der Dimension Schreiben (halboffen Schreiben einfach) kein Effekt auf¹¹⁵.

Zusammenfassend wurden bei 9 von 31 Aufgaben und 4 von 8 Formaten signifikante Werte für einen möglichen Zusammenhang zwischen der Aufgabebearbeitungszeit und der Computererfahrung identifiziert.

Der Hypothese $H1_{a2}$ kann zwar somit nur sehr bedingt zugestimmt und die Hypothese $H0_{a2}$ kann nicht verworfen werden. Jedoch ist unter Berücksichtigung der Formate als Ergebnis festzuhalten, dass Personen mit wenig Computererfahrung mehr Zeit für die Bearbeitung der offenen, halboffenen und Korrekturformate benötigen als Personen mit viel Computererfahrung.

¹¹³ Auch hier gelten die Ergebnisse der auf der Power-Analyse und somit das Signifikanzniveau von 0,1.

¹¹⁴ Es werden nur die „formatreinen“ Aufgaben in den Berechnungen berücksichtigt. Die Aufgaben 2.1.02_I, 2.1.02_II, 2.4.01_I und 2.4.01_II sind daher in die Berechnungen nicht einbezogen.

¹¹⁵ Das Format „Single Choice Lesen schwer“ wird aufgrund zu geringer Fallzahlen nicht berücksichtigt.

Ein Zusammenhang zwischen der Computererfahrung und der formatunabhängigen Aufgabenbearbeitungszeit kann aber nicht bestätigt werden.

Bearbeitung der Frage F2b: „Haben die Computererfahrung und -kompetenz einen Einfluss auf die Performanz?“

Um die Frage F2b zu prüfen, wird ein Mittelwertvergleich zwischen der Computererfahrung und der Performanz mit einem t-Test für unabhängige Stichproben durchgeführt¹¹⁶. Für die Computererfahrung werden die Kategorien zur Computererfahrung „viel Erfahrung“ sowie „wenig Erfahrung“ und für die Performanz die relativen Performanzwerte herangezogen (vgl. auch Kap. 8.1.2.2 zu den methodischen Vorüberlegungen).

Die Mittelwertvergleiche ergeben statistisch signifikante Unterschiede zwischen der Computererfahrung und des relativen Performanzwerts des ersten Durchlaufs ($p=.02$) sowie für beide Durchläufe zusammen ($p=.04$) jedoch nicht separat für den zweiten Durchlauf¹¹⁷.

Schlussfolgernd kann die Hypothese H_{0b} nicht verworfen und der Hypothese H_{1c} nur bedingt zugestimmt werden. Jedoch zeichnet sich eine Tendenz hinsichtlich der H_{1c} ab: Es scheint ein Zusammenhang zwischen der Computererfahrung und der Performanz zu bestehen. Lediglich für den separaten zweiten Durchlauf und die formatbezogenen Performanzwerte ergeben sich keine signifikanten Werte, was in der geringeren Fallzahl begründet sein kann. Mit der Bearbeitung der Frage F2c wird im Folgenden exploriert, worin diese Ergebnisse begründet sein können.

Bearbeitung der Frage F2c: „Was muss bei der Usability berücksichtigt werden, damit Computererfahrung und -kompetenz keinen Einfluss auf das Testergebnis haben?“

Um zu untersuchen, was bei der Usability zu berücksichtigen ist, damit die Computererfahrung und -kompetenz keinen Einfluss auf das Testergebnis haben ist zu explorieren, worin der tendenzielle Effekt der Computererfahrung auf die relative Performanz begründet ist. Hierfür werden vorerst die durchschnittlichen

¹¹⁶ Die Normalverteilung ist lediglich bei dem relativen Performanzwert für das Format „Single Choice Lesen schwer“ nicht gegeben, was vermutlich in geringen Fallzahl zu begründen ist ($n=4$). Dieses Format wird bei der Analyse der Ergebnisse somit nicht berücksichtigt.

¹¹⁷ Vermutlich sind für den zweiten Durchlauf die Fallzahlen zu gering, um aussagekräftige Daten zu generieren.

Performanzwerte dargestellt – aufgeschlüsselt nach Aufgabenformat sowie der Computererfahrung:

	Computererfahrung	N	MW	SD	Standardfehler des MW
Single Choice Lesen einfach Durchlauf 1	wenig	7	70%	27%	10%
	viel	3	68%	13%	7%
Single Choice Lesen mittel Durchlauf 1	wenig	7	63%	19%	7%
	viel	6	72%	22%	9%
Single Choice Lesen schwer Durchlauf 1	wenig	1	24%	,	,
	viel	0 ^a	,	,	,
halboffen Schreiben einfach Durchlauf 1	wenig	12	50%	25%	7%
	viel	10	57%	14%	4%
offen Schreiben einfach Durchlauf 1	wenig	5	54%	34%	15%
	viel	6	80%	25%	10%
halboffen Schreiben mittel Durchlauf 1	wenig	7	60%	24%	9%
	viel	6	71%	22%	9%
Korrektur Schreiben mittel Durchlauf 1	wenig	7	33%	47%	18%
	viel	6	72%	33%	13%
Offen Schreiben mittel Durchlauf 1	wenig	7	38%	31%	12%
	viel	6	52%	26%	11%

a. T kann nicht berechnet werden, da mindestens eine der Gruppen leer ist.

Tabelle 39: Computererfahrung und relative formatbezogene Performanzwerte für Durchlauf 1

Für den ersten Durchlauf sind die Mittelwerte für die relative Performanz fast bei allen Personen mit wenig Computererfahrung niedriger als bei Personen mit hoher Computererfahrung. Eine Ausnahme bildet lediglich das Format Single Choice im Bereich Lesen auf einfachem Niveau.

Für den ersten und zweiten Durchlauf zusammen sind ebenfalls die relativen Performanzwerte im Durchschnitt für Personen mit viel Computererfahrung höher, jedoch sind die Ergebnisse hier nicht so konsistent wie für den ersten Durchlauf:

	Computererfahrung	N	MW	SD	Standardfehler des MW
Single Choice Lesen einfach Durchlauf 1	wenig	10	76%	25%	8%
	viel	4	69%	10%	5%
Single Choice Lesen mittel Durchlauf 1	wenig	7	63%	19%	7%
	viel	13	78%	20%	6%
Single Choice Lesen schwer Durchlauf 1	wenig	2	48%	34%	24%
	viel	1	43%	,	,
halboffen Schreiben einfach Durchlauf 1	wenig	12	50%	25%	7%
	viel	10	57%	14%	4%
offen Schreiben einfach Durchlauf 1	wenig	10	70%	29%	9%
	viel	9	69%	32%	11%
halboffen Schreiben mittel Durchlauf 1	wenig	10	55%	23%	7%
	viel	12	58%	25%	7%
Korrektur Schreiben mittel Durchlauf 1	wenig	10	47%	48%	15%
	viel	11	82%	27%	8%
Offen Schreiben mittel Durchlauf 1	wenig	10	37%	31%	10%
	viel	11	55%	22%	7%

Tabelle 40: Computererfahrung und relative formatbezogene Performanzwerte für Durchlauf 1 und 2

Bei den Formaten *offen Schreiben einfach*, *Single Choice Lesen einfach* und *Single Choice Lesen schwer* sind die durchschnittlichen relativen Performanzwerte allerdings höher für Personen mit wenig Computererfahrung. Im Vergleich zu den Daten des ersten Durchlaufs zeigt sich hier ein heterogenes Bild.

Fraglich ist, ob bestimmte Formate identifiziert werden können, bei denen die Computererfahrung einen Einfluss auf die relative Performanz hat. Die Berechnungen für Computererfahrung und den ersten Durchlauf, für Computererfahrung und den zweiten Durchlauf sowie für Computererfahrung und beide Durchläufe zusammen ergeben lediglich für das Format *Korrektur Schreiben mittel* für beide Durchläufe zusammen einen signifikanten Unterschied ($p=.06$).

Insgesamt wird deutlich, dass insbesondere für den ersten Durchlauf für die Items auf den schwierigsten Niveaus (Lesen schwer und Schreiben mittel) tendenziell ein niedrigerer Performanzwert erzielt wurde als auf den einfacheren Niveaus. Es kann allerdings kein statistischer Zusammenhang festgestellt werden.

Zusammenfassend ist für den Fragebereich zur Computererfahrung und Computerkompetenz festzuhalten, dass ein Zusammenhang zwischen der Computererfahrung und der eingeforderten Unterstützung besteht: Probanden und Probandinnen mit wenig Computererfahrungen werden öfter unterstützt als Probanden und Probandinnen mit viel Computererfahrung. Ein systematischer Zu-

sammenhang zwischen der Computererfahrung und der formatunabhängigen Aufgabenbearbeitungszeit kann nicht bestätigt werden, es zeichnet sich allerdings für die Formate auf höherem Niveau und tendenziell offene Formate ein Zusammenhang dahingehend ab, dass Personen mit wenig Computererfahrung mehr Zeit für die Bearbeitung der Aufgaben benötigen als Personen mit viel Computererfahrung. Zwischen der Computererfahrung und der Performanz kann ebenfalls kein systematischer Zusammenhang festgestellt werden. Allerdings zeigen die Werte eine Tendenz dahingehend, dass die Computererfahrung einen Einfluss auf die Performanz hat.

8.3.6 Diskussion der Ergebnisse zur Computererfahrung und -kompetenz

Die Computererfahrung hat insbesondere einen Einfluss auf die Aufgabenbearbeitungszeit der halboffenen und offenen Formate.

Begründet ist dies vermutlich darin, dass für die genannten Formate sowohl stärkere Konstruktionsleistungen (z. B. das Buchstabieren eines Wortes) als auch die aktive Produktion der Antwort in Form von Eintippen erbracht werden müssen. Bei geschlossenen Formaten werden den Probanden und Probandinnen drei Antworten zur Auswahl gestellt. Die Identifikation der richtigen Antwort bedarf keiner eigenständigen Konstruktionsleistung. Personen mit wenig Computererfahrung benötigen vermutlich mehr Zeit, um die Buchstaben für das Eintippen auf der Tastatur zu finden. Dieses Ergebnis entspricht zudem dem Ergebnis in der Studie von Bennett u. a. (2008), in der herausgefunden wurde, dass der Effekt auf das Testergebnis umso größer war, je offener das Antwortformat war (vgl. auch Kap. 2.2.6 zu innovativen Itemformaten).

Des Weiteren kann die Auswahl der richtigen Antwort durch Distraktoren, die sich stark von der richtigen Antwort unterscheiden, unterstützt werden. Da die Formate, bei denen ein Effekt zwischen Bearbeitungszeit und Computererfahrung festgestellt wurde, sowohl dem leichten als auch dem schweren Niveau angehören, ist ein Einfluss der Literalitätskompetenz nicht anzunehmen¹¹⁸. Somit können zwei Vermutungen angestellt werden:

- 1) Die Bearbeitungszeit ist bei den halboffenen, offenen und Korrekturformaten höher, weil die Erschließung und Bearbeitung der Formate einen höheren *Extraneous Load* beansprucht als Single Choice-Formate.

¹¹⁸ Ein Effekt kann jedoch nicht vollkommen ausgeschlossen werden: Personen, die geringe literale Kompetenzen haben, arbeiten auch wenig mit Computern (vgl. Kap. 2.3 zu PIAAC). Somit benötigen diese Personen aufgrund ihrer niedrigeren Literalität im Vergleich zu Personen mit viel Computererfahrung und höherer Literalitätskompetenzen mehr Zeit für die Bearbeitung der Aufgaben auf höherem Niveau. Auf niedrigem Niveau wäre der Effekt nicht sichtbar, da die Diskriminierung der Items auf niedrigem Niveau ggf. zu gering ist.

- 2) Die Bearbeitungszeit ist bei den halboffenen, offenen und Korrekturformaten höher, weil die Formate eine eigenständige Konstruktionsleistung erfordern, wohingegen Single Choice-Formate lediglich der Identifikation der richtigen Antwort bedarf.

Die Vermutungen schließen sich nicht gegenseitig aus. Es ist möglich, dass sowohl die stärkere Beanspruchung des *Extraneous Load* als auch die geforderte Konstruktionsleistung zu einer stärkeren Beanspruchung des *Cognitive Load* insgesamt führen und daher mehr Zeit für die Bearbeitung der Aufgaben benötigt wird.

Bezüglich der Überprüfung eines möglichen Zusammenhangs zwischen der Computererfahrung und der Performanz kann kein statistischer Zusammenhang festgestellt werden. Dieses Ergebnis legitimiert zudem, dass otu.lea gleichzeitig Evaluations- als auch Erhebungsinstrument ist. Die Tendenz hin zu einer Korrelation zwischen der Computererfahrung und der Performanz kann darin begründet sein, dass Personen mit wenig Computererfahrung auch tendenziell niedrigere Lese- und Schreibkompetenzen vorweisen. Auf einen kausalen Zusammenhang deutet auch die aktuelle PIAAC-Studie hin: Wie bereits in Kap. 4.2.1 dargestellt, wurde in der PIAAC-Studie 2012 untersucht „wie Personen sich Informationen in einer computergestützten Umgebung erfolgreich beschaffen und wie sie diese verwenden. Hierzu wurden Aufgaben wie das Sortieren und Versenden von E-Mails, die Bearbeitung von virtuellen Formularen sowie die Beurteilung des Informationsgehalts und der Vertrauenswürdigkeit verschiedener Internetseiten eingesetzt.“ (Rammstedt u. a., 2013, S. 12). Die Ergebnisse zeigen, dass Personen ohne Computererfahrung durchschnittlich niedrigere Werte in den Bereichen Lesekompetenz und alltagsmathematische Kompetenzen erreichen als Personen, bei denen die Kompetenzmessung computergestützt durchgeführt werden konnte (Zabal u. a., 2013, S. 68 ff.). Personen mit geringer Lesekompetenz scheinen somit auch weniger Computererfahrung haben (für eine genaue Erläuterung vgl. Kap.4.2.1).

Es besteht offensichtlich ein gewisser Zusammenhang zwischen Lesekompetenz, alltagsmathematischer Kompetenzen und der Computererfahrung:

„Damit neue Technologien kompetent und zweckdienlich zur Lösung von alltäglichen Problemen, zum Beispiel zur Informationsbeschaffung, eingesetzt werden können, sind nicht nur die grundlegenden technische Handhabung von Hard- und Software von Bedeutung, sondern insbesondere auch kognitive Fähigkeiten, wie Lese-, mathematische und Problemlösekompetenz wichtig (International ICT Lite-

racy Panel, 2007)“ (zitiert nach Zabala u. a., 2013, S. 60).

Kausal lässt sich dieser Zusammenhang womöglich dadurch begründen, dass die Nutzung von Computern eine minimale Lese- und Schreibkompetenz voraussetzt und daher Personen mit niedrigen Lese- und Schreibkompetenzen keine Computer nutzen und somit auch den Umgang damit nicht erlernen. Das deutet wiederum hinsichtlich der hier diskutierten Ergebnisse darauf hin, dass Personen mit wenig Computererfahrung auch einen niedrigeren Performanzwert erzielen als Personen mit einer ausgeprägten Computererfahrung. Es kann allerdings nur gemutmaßt werden, ob sich bei einer Untersuchungsdurchführung mit einer größeren Stichprobe mit *otulea* ein Zusammenhang zeigt, der im Vergleich zu einem möglichen Zusammenhang zwischen der Computererfahrung und der Performanzwerte geschlossener Itemformate auf niedrigerem Niveau stärker ist.

Für weitere Untersuchungen zu einem möglichen Einfluss der Computererfahrung auf die Performanz könnten weitere Validitätsüberprüfung, mit Hilfe eines Vergleichs der Ergebnisse mit dem papierbasierten sowie dem computerbasierten Instrument, durchgeführt werden. Da in unterschiedlichen Aufgaben mitunter die gleichen Kannbeschreibungen geprüft werden, ist der Vergleich beiden Erhebungsformen mit verschiedenen Items möglich. Das würde bedeuten, es werden mit zwei unterschiedlichen Erhebungsformen (papier- und computerbasiert) die gleichen Kompetenzen gemessen; durch die Nutzung verschiedener Items, die aber das gleiche messen, kann ein potenzieller Lerneffekt ausgeschlossen werden. Mit diesem Setting könnte zudem die Äquivalenzproblematik thematisiert und überprüft werden.

Eine weitere Überlegung hinsichtlich der Frage, was bei der Usability berücksichtigt werden muss, damit die Computererfahrung keinen Einfluss auf die Performanz hat, ist die Einbindung einer *adaptiven* Hilfefunktion. Wie auch die Ergebnisse zur Selbstbeschreibungsfähigkeit (vgl. Kap. 8.3.3) zeigen, kann auf der Grundlage der Ergebnisse vorerst nicht davon ausgegangen werden, dass die Nutzung der Hilfefunktion zum gewünschten Effekt führt. Dass die Computererfahrung tendenzielle einen Einfluss auf die Performanz hat, kann auch in einem mangelnden Effekt der Hilfefunktion begründet sein. Zu untersuchen wäre, ob eine adaptive Hilfefunktion einen positiven Effekt hervorruft – nämlich die Bearbeitung der Aufgaben ohne, dass die Computererfahrung einen Einfluss auf das Testergebnis ausüben würde. Eine adaptive Hilfefunktion würde nicht nur kontextsensitiv gestaltet sein, sondern sie würde entsprechend des Niveaus und der Aktionen, bis zum Zeitpunkt des Aufrufs der Hilfefunktion, angepasst werden. Wenn beispielsweise eine Person mehrfach die Hilfefunktion genutzt hat kann das ein Hinweis darauf sein, dass die Hilfefunktion nicht die Inhalte anbie-

tet, die zum Bearbeiten der Aufgabe notwendig sind. Von der Anzahl und des Kontextes des Abrufs abhängig, könnten entweder weitere Hinweise zur Bearbeitung der jeweiligen Seite angeboten werden. Oder die Inhalte der bereits abgerufenen Hilfefunktion könnten anders aufbereitet dargestellt werden.

Darüber hinaus stützen die Ergebnisse zum Zusammenhang zwischen der Computererfahrung und der Performanz des ersten Durchlaufs die Überlegungen zur zusätzlichen Heuristik „Übungssequenz“ (vgl. Kap. 8.3.3 zu Überlegungen für zusätzliche Heuristiken). Hinsichtlich der Validität der Testumgebung ist zu gewährleisten, dass die Computererfahrung keinen Einfluss auf die Performanz ausübt. Bei einer Vorschaltung von Übungssequenzen kann eine mangelnde Computererfahrung kompensiert werden, indem die Teilnehmenden sich mit den grundsätzlichen Funktionen zur Nutzung der Testumgebung vertraut machen können.

8.3.7 *Darstellung der Ergebnisse zum Extraneous Load*

Bearbeitung der Frage F3a: „Gibt es Hinweise auf die Beanspruchung des Extraneous Load durch den Einsatz von Cues?“

Mit Hilfe objektiver Messmethoden wurden in der experimentellen Usability-Studie Ergebnisse zur Performanz und Bearbeitungszeit generiert, die Hinweise auf die Auslastung des Cognitive Load liefern sollen.

Die Cues waren bei 25 Personen aktiviert und entsprechend bei 25 Personen nicht aktiviert. Um zu prüfen, ob die Cues einen Effekt auf die Performanz haben, wird mit den relativen Performanzwerten ein t-Test für unabhängige Stichproben durchgeführt¹¹⁹. Einbezogen werden die relativen Performanzwerte.

Die Berechnungen ergeben bei einem Signifikanzniveau von 0,1¹²⁰ für den ersten ($p=.07$) und zweiten Durchlauf ($p=.08$) signifikante Ergebnisse. So scheint der Einsatz von Cues einen Effekt auf die Performanz zu haben. Der Einsatz von Cues hat bei dem ersten und dem zweiten Durchlauf allerdings einen entgegengesetzten Effekt: Beim ersten Durchlauf hat der Einsatz von Cues einen negativen Effekt auf die relative Performanz: Werden Cues eingesetzt, ist der Performanzwert tendenziell niedriger. Beim zweiten Durchlauf hat der Einsatz von Cues einen positiven Effekt auf die Performanz: Werden Cues eingesetzt, ist die Performanz tendenziell höher. Folglich ergeben sich unter Berücksichtigung beider Durchläufe keine signifikanten Werte.

¹¹⁹ Die Bedingungen für die Durchführung eines t-Tests sind erfüllt.

¹²⁰ Bei einer angenommenen großen Effektstärke von 0,8 (vgl. z. B. Bortz & Döring, 2006, S. 642) und einem angenommenen Alpha-Fehler von 0,1 ergibt sich eine Power von 0,9. Somit wird das Signifikanzniveau auf 0,1 festgelegt.

Für weitere Erklärungen dieser Effekte wird exploriert, ob der Effekt bei bestimmten Itemformaten (stärker) auftritt. Für den ersten Durchlauf ergibt der t-Test für unabhängige Stichproben signifikante Werte für die Formate *Korrektur Schreiben mittel* ($p=.02$) und *offen Schreiben mittel* ($p=.03$). Die Verteilung wird durch Boxplots veranschaulicht:

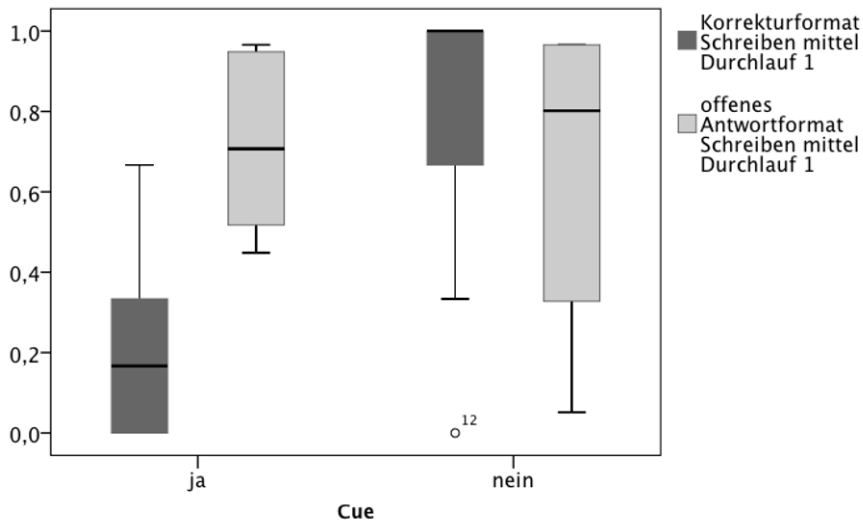


Abbildung 37: Boxplot für Durchlauf 1: Cue und ausgewählte Aufgabenformate

Beim Korrekturformat ist ein besonders großer Unterschied festzustellen. Bei beiden Formaten hat der Einsatz von Cues einen negativen Effekt auf die Performanz.

Für den zweiten Durchlauf sind ebenfalls signifikante Werte für die Formate *Korrektur Schreiben mittel* ($p=.06$) und *offen Schreiben mittel* ($p=.07$) festzustellen, deren Verteilung die folgende Abbildung zeigt:

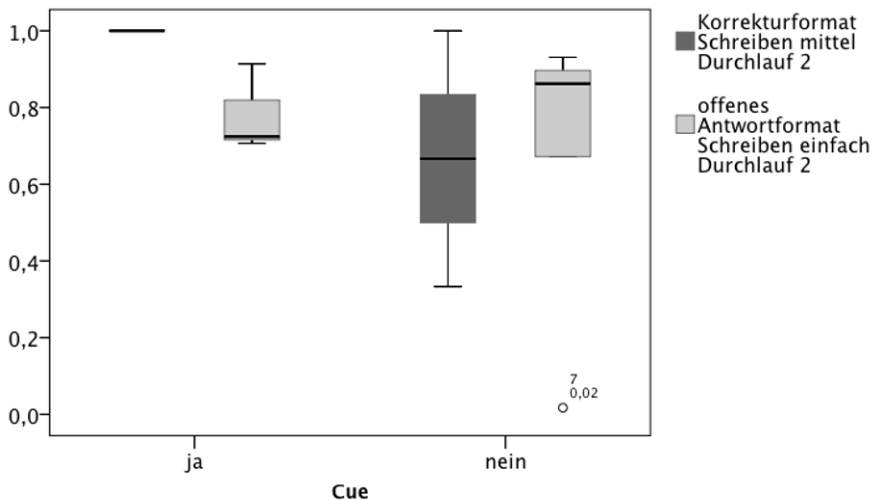


Abbildung 38: Boxplot für Durchlauf 2: Cue und ausgewählte Aufgabenformate

Für den zweiten Durchlauf zeigt sich ein kontroverses Bild: Für das Korrekturformat scheint der Einsatz von Cues einen positiven Effekt zu haben (werden Cues eingesetzt, sind die Performanzwerte tendenziell höher), für das offene Antwortformat einen negativen Effekt (der Einsatz von Cues hat einen tendenziell niedrigeren Performanzwert zur Folge). Der dennoch signifikante Wert für den zweiten Durchlauf und der positive Effekt bei formatunabhängiger Betrachtung können durch den relativ großen Unterschied beim Korrekturformat hervorgerufen worden sein. Da die Ergebnisse ein kontroverses Bild hinsichtlich eines möglichen Effekts von Cues zeigen, bleibt die Nullhypothese H_{0c1} bestehen:

Es besteht kein Zusammenhang zwischen dem Einsatz von Cues und der Performanz.

Für die Überprüfung, ob die Cues einen Effekt auf die Bearbeitungszeit haben, wird auf die Berechnung zur Wahrnehmungssteuerung zurückgegriffen (s. Kap. 8.3.3): Die Ergebnisse zeigen, dass der Einsatz von Cues lediglich einen Effekt auf die Bearbeitungszeit der Aufgabe 1.2.1 hat.

Daher kann auch die H_{0c2} nicht verworfen werden und bleibt bestehen.

Es besteht kein Zusammenhang zwischen dem Einsatz von Cues und der Aufgabebearbeitungszeit.

Bearbeitung der Frage F3b: „Gibt es weitere Hinweise auf Faktoren, die den Cognitive Load im Allgemeinen und den Extraneous Load im Besonderen beeinflussen?“

Weitere potenzielle Auslastungsfaktoren für den *Cognitive Load* sind a) das Vorwissen und b) die subjektiv empfundene Schwierigkeit der Nutzung von otu.lea.

a) Vorwissen

Das Vorwissen wird über die Computererfahrung erhoben. Die subjektive Einschätzung wird über die Interviews sowie den UEQ beurteilt.

Um das Vorwissen als mögliche Einflussgröße zu berücksichtigen, ist die Verteilung der Gesamtstichprobe auf beide Gruppen (Cue und Computererfahrung) zu prüfen. Hinsichtlich der Computererfahrung (das Vorwissen) ist festzustellen, dass eine relativ gleichmäßige Verteilung besteht:

	Computererfahrung			Gesamtsumme
	wenig		viel	
Cue	ja	14	10	24
	nein	13	12	25
Gesamtsumme	27		22	49

Tabelle 41: Verteilung zwischen experimenteller Bedingungen (Cue) und Computererfahrung

Die Tabelle zeigt, dass eine recht gleichmäßige Verteilung vorliegt. Es kann daher davon ausgegangen werden, dass die Ergebnisse nicht durch die Computererfahrung moderiert werden. Zudem ist aufgrund der Selbsteinschätzung anzunehmen, dass die Probanden und Probandinnen ein Niveau gewählt haben, das tendenziell ihrem Kompetenzstand entspricht. Somit kann davon ausgegangen werden, dass die Beanspruchung des Intrinsic Load aufgrund der Aufgabenschwierigkeit bei den Probanden und Probandinnen nicht stark voneinander abweicht. Folglich ist anzunehmen, dass das Vorwissen aufgrund der Computer-

erfahrung keinen moderierenden Einflussfaktor auf die Beanspruchung des *Extraneous Load* darstellt.

b) Subjektiv empfundene Schwierigkeit der Nutzung von otu.lea

Ergänzend zu den Daten der objektiven Messmethoden¹²¹ werden im Folgenden Interviewpassagen aufgeführt, die Hinweise auf die Beanspruchung des *Extraneous Load* liefern.

Eine Person trifft eine Aussage über die Schwierigkeit, sich die Aufgabe zu erschließen¹²²:

I: „Mhm. Und was ist Ihnen in Bezug auf die Schrift aufgefallen?“

C3: „Ist gut. Es wäre gut wenn unterschiedliche Farben zwischen der Frage und der Aufgabe verwendet werden würden.“

I: „Wir müssen mal gucken, ich hab hier ein paar Screenshots.“

C3: „Ja genau das ist die Frage.“

I: „Ja.“

C3: „Das komm an Anfang [...] wäre gut wenn die Farbe ein bisschen anders wäre.“

I: „Wenn sich die Farbe von der Frage abheben würde zu den Antworten?“

C3: „Genau, genau, das würde weiterhelfen, besser zu begreifen.“

I: „Ah ja.“

C3: „Weil wenn du nicht gut in deutscher Sprache bist...“

I: [...]

C3: „...kann das verwirren.“

I: „Ja.“

C3: „Man muss viel an anderes denken (vermutlich ist hier die Aufgabenbearbeitung gemeint) und wenn man zusätzlich über die Farbe nachdenken muss (unverständlich).“

¹²¹ Die Skala „Durchschaubarkeit“ des UEQ kann aufgrund einer niedrigen Skalenkonsistenz (s. Kap. 8.3.2.3) nicht zu weiteren Berechnungen herangezogen werden.

¹²² Der Interviewausschnitt ist stark paraphrasiert, da es sich bei der Interviewpartnerin/dem Interviewpartner um eine Person mit niedrigen Deutschkenntnissen handelt.

Diese Aussage liefert einen Hinweis auf Ursachen für die Beanspruchung des *Extraneous Load*, indem die Person die Aufgabe nicht inhaltlich, sondern in ihrer farblichen Darstellung als schwierig empfand. Er/Sie berichtet darüber, dass die Einheiten der Aufgaben (Frage und Aufgabe) farblich stärker voneinander abgegrenzt werden sollen, um sich die Aufgabe besser erschließen zu können und um von der farblichen Trennung auch auf eine inhaltliche Trennung schließen zu können. Die Person geht davon aus, dass sie damit die Aufgabe besser verstehen würde. Da sich die Aussage auf die Gestaltung und Darbietung der Informationen bezieht, ist diese dem *Extraneous Load* und speziell der Aufnahme von visuellen Informationen zuzuordnen. Um die Beanspruchung des *Extraneous Load* zu reduzieren, ist zu überlegen, das Gesetz der Nähe (s. Kap. 2.6.3 zu kognitiven und physiologischen Voraussetzungen) stärker zu berücksichtigen und Seitenbestandteile, die inhaltlich voneinander abgrenzbar sind farblich stärker zu kontrastieren.

Einen Hinweis für einen reduzierenden Faktor der Auslastung des *Extraneous Load* ist im folgenden Zitat beschrieben:

I: „Okay mhm. Ähm und gab es denn irgendwas was dir in dem Programm geholfen hat. Was dir das ganze erleichtert hat? Der Umgang. Den Umgang mit dem Programm.“

E3: „Ja, dass da immer stand was man machen muss. Start klicken oder so.“

Die Aussage dieser Person erachtet die Instruktionen zur Interaktion (klicken, tippen etc.) als unterstützend. Auch die folgende empfindet die Instruktion als unterstützend. Sie beschreibt, dass die Instruktionen dazu beitragen, die relevanten Aspekte fokussieren zu können:

I: „Fällt Ihnen etwas zu dem Aufbau ein, also wie die Seiten aufgebaut waren, das Programm, können Sie dazu irgendetwas sagen? Was Ihnen dazu einfällt?“

O3: „Das Gute ist, wenn man anfängt, dass man oben drauf drückt und man weiß schon mal was sie hinterher von einem will. Sie liest den Text komplett vor, wenn man auf den Dings geht und dann weiß ich schon aha, pass auf. Mit der Putzfrau zum Beispiel. Da fehlen Sachen, ich wusste genau, worauf ich achten musste, dass muss irgendwo was fehlen,

also muss ich Buchstaben einsetzen. Also doch, war gut.“

Beide Personen berichten damit, dass die von der Testumgebung bereitgestellten – nicht auf die inhaltliche Bearbeitung bezogenen – Informationen sie bei der Aufgabenbearbeitung unterstützen. Dies lässt die Vermutung zu, dass die Instruktionen zu einem reduzierenden Effekt der Beanspruchung des *Extraneous Load* führen.

Auch die folgende Aussage ist ein Indiz für eine tendenziell niedrige Beanspruchung des *Extraneous Load*:

I: „Also, dass du sagen könntest, dass fandst du jetzt konkret gut, dass es so und so aufgebaut war?“

J3: „Ja, dass es alles so schön hintereinander geht. Das man erst alles... erstmal die Frage, die sie dann gesagt haben. Dann darauf dann halt der Text und dann darauf das, was du dann z. B. bei dem Schreiben noch an der Seite schreiben kannst oder musst.“

I: „Ja.“

J3: „Und das du dann auch noch wieder extra die Wörter dir anhören kannst.“

Hier wird positiv der iterative Seitenaufbau betont. Die Aussage lässt vermuten, dass der iterative Aufbau dazu führt, sich nacheinander auf die Aufgabensequenzen konzentrieren zu können und nicht die Gefahr eines Overload besteht.

Die Probanden und Probandinnen bewerten otu.lea nicht einheitlich hinsichtlich der Informationsfülle und des Schwierigkeitsgrades.

V3: „Man muss sehen, man kanns, aber ich finde es halt etwas merkwürdig weil das so leicht war, man weiß halt nicht wie man sich fühlen soll.“

Anzumerken ist, dass diese Aussage vermutlich auf den Inhalt der Aufgaben bezogen ist. Dennoch ist das ein Hinweis auf eine mögliche *Oversimplification* (vgl. Kap. zu Cognitive Load), die zu Irritationen, Frustration und Motivationsenkung führen kann. Ein weiterer Hinweis auf eine *Oversimplification* ist die Aussage einer Person, das notwendige wiederholte klicken auf den Lautsprecher, um fortzufahren, sei nervig (T3, Z. 10-10, vgl. auch Kap. 8.3.3 Ergebnisdarstellung zur Heuristik „Steuerbarkeit“).

Von einer anderen Person wird angemerkt, dass die Informationsfülle teilweise zu hoch ist:

I: „Und ähm, inhaltlich?“
W3: „Inhaltlich auch gut, bloß, manchmal (hab ich das schwierig, das zu verstehen?) [...]“.“
I: „Dass sie zu viel?“
[...]
W3: „Dann kommt zu viel und dann wird das für mich zu viel dann.“
I: „Mhm.“
W3: „Und dann hab ich das manchmal durcheinander.“
I: „Wie bitte?“
W3: „Dann hab ich das manchmal durcheinander.“

Hier wird deutlich, dass sequentiell zu viele Informationen vorhanden sind. Es wird allerdings nicht erläutert, ob es sich um aufgaben- oder formatbezogene Informationen handelt. Es wird aber deutlich, dass es bei dieser Person teilweise zu einer *Overcomplication* kommt.

Einen weiteren Hinweis zur Auslastung des Arbeitsgedächtnisses liefert eine Person, indem sie die Möglichkeit, sich die Instruktionen und Wörter bzw. Sätze mehrmals anhören zu können, positiv betont:

J3: „Weil manchmal, wenn du dann so lange Texte hast, dann hast du manchmal irgendwie so ‘n kleinen... Dreher, dass du’s dann nicht mehr weißt. Und deswegen fand ich das dann halt gut, dass man die dann wieder anklicken konnte.“

Die Person beschreibt, dass es mitunter schwierig ist, die gesamte Instruktion bzw. die ganzen Sätze im Kopf zu behalten. Das deutet auf die Begrenztheit des Arbeitsgedächtnisses und in diesem Fall des *Intrinsic Load* hin. Die Möglichkeit, sich mehrmals etwas vorlesen lassen zu können, reduziert die Informationen, die im Arbeitsspeicher vorgehalten werden müssen und somit auch die Beanspruchung der kognitiven Kapazität, so dass mehr Ressourcen für die weitere Bearbeitung der Aufgabe vorhanden sind. Dieser Effekt wird auch in *Limited Capacity Assumption* (der Annahme über der Begrenztheit des Arbeitsgedächtnisses) (vgl. Mayer 2005, 35 und Kap. 2.4) beschrieben.

Auch die folgende Person beschreibt, wie die Darstellungen in der Testumgebung das Verständnis unterstützen:

I: „Ja? Was is‘ dir so besonders dran aufgefallen?
Oder kann auch nichts sein, also kann beides sein.“

AJ3: „Diese Bild passte immer zum Text.“

I: „Okay.“

AJ3: „Und dann kann man das besser verstehen.“

Die Aussage bestätigt die im Theorieteil dargelegte *Dual Channel Assumption* (die Annahme darüber, dass Informationen über zwei Kanäle wahrgenommen und verarbeitet werden) (vgl. Mayer 2005, 33 und Kap. 2.4.2 zu CTML) und diese besser verarbeitet werden. In der Aussage dieser Person wird nicht deutlich, ob sie den geschriebenen oder den gesprochenen Text meint. Falls ihre Aussage auf den gesprochenen Text bezogen ist, kann von dieser Aussage auf das Modalitätsprinzip geschlossen werden. Dieses besagt, Inhalte können besser verarbeitet und behalten werden, wenn Bild und Ton miteinander kombiniert werden (vgl. Low & Sweller 2005 und vgl. Kategorie „Inhaltsdarstellung“ im Kap. 2.2.6 zu innovativen Itemformaten).

Die Ergebnisse des UEQ zu den Gegensatzpaaren „verständlich–unverständlich“, „leicht–schwer“, „kompliziert–einfach“ können nicht berücksichtigt werden. Aufgrund einer zu niedrigen Skalenkonsistenz (ausgegeben durch Cronbachs α) ist davon auszugehen, dass die Gegensatzpaare nicht einheitlich interpretiert wurden. Es ist möglich, dass die Gegensatzpaare nicht auf die Usability von otu.lea, sondern auf die inhaltliche Aufgabenschwierigkeit bezogen wurden.

8.3.8 Diskussion der Ergebnisse zum Extraneous Load

Nachfolgend werden die Ergebnisse zu den potenziellen Auslastungsfaktoren diskutiert. Es wird sich dabei an der Struktur des vorherigen Kapitels orientiert, indem vorerst die beobachteten Effekte von Cues und nachfolgend weitere Hinweise auf Auslastungsfaktoren des Cognitive Load (durch Angaben zum Vorwissen sowie zu subjektiven Einschätzungen) diskutiert werden.

Diskussion der beobachteten Effekte durch den Einsatz von Cues

Hinsichtlich der Performanz kann unter Berücksichtigung der Ergebnisse des ersten und zweiten Durchlaufs der Effekt darin begründet sein, dass die Cues beim ersten Durchlauf für das Korrekturformat noch unbekannt sind und den *Extraneous Load* stärker beanspruchen, wohingegen die Cues beim zweiten Durchlauf keine starke Beanspruchung des *Extraneous Load* nach sich ziehen und damit die Bearbeitung unterstützen. Dieser Effekt lässt sich folgendermaßen erklären: Die Cues werden im ersten Durchlauf von den Probanden und Probandinnen als zusätzliche Information wahrgenommen. Diese Information wird kognitiv verarbeitet und beansprucht daher kognitive Ressourcen. Folglich steigt der *Extraneous Load* und mit ihm auch der gesamte *Cognitive Load* vorerst an

(vgl. Sweller, 2010, S. 44 und Kap. 2.4.1). Im zweiten Durchlauf wurden die Informationen verarbeitet, indem die Funktion der Cues gelernt und erfolgreich als mentales Modell integriert wurden. Die Cues werden somit im zweiten Durchlauf als Unterstützung wahrgenommen, erleichtern die Orientierung und haben daher einen reduzierenden Effekt auf die Beanspruchung des *Extraneous Load*. Dies führt wiederum dazu, dass mehr kognitive Ressourcen für die inhaltliche Bearbeitung vorhanden sind.

Für das offene Format hat der Einsatz von Cues sowohl beim ersten als auch beim zweiten Durchlauf einen negativen Effekt. Daher ist hier die Frage, ob die Cues ggf. dauerhaft keinen unterstützenden Charakter haben. Die offenen Formate bestehen vorwiegend aus Lückentexten. Sobald der Aufgabenbereich erscheint, blinkt der erste Lautsprecherbutton, durch dessen Aktivierung das Wort/die Wörter vorgelesen wird/werden, das/die in die Lücke eingetragen werden soll/sollen. Anschließend wird das Textfeld animiert. Nachdem etwas in das Textfeld eingetragen wurde, blinkt in chronologischer Reihenfolge des Textes der nächste Lautsprecherbutton. Der Lautsprecherbutton fängt zu blinken an, sobald wenige Sekunden kein Eintrag in das Textfeld erfolgt. Die Reihenfolge zur Bearbeitung des Lückentextes wird somit zwar vorgegeben, doch ist die Animation nicht abhängig von dem inhaltlichen Eintrag in ein Textfeld. Wenn eine Person die Eingabe in ein Textfeld unterbricht, um über die Lösung nachzudenken, der darauf folgende Lautsprecher aber bereits blinkt, kann das zu Irritationen führen oder auch ablenkend wirken.

Dass weder hinsichtlich der Performanz noch der Aufgabenbearbeitungszeit systematischen Effekte beobachtbar sind, kann zudem darin begründet sein, dass die Beanspruchung des *Intrinsic Load* zu niedrig war, um einen Effekt auf den *Extraneous Load* beobachten zu können. Wie im theoretischen Teil zur CLT dargelegt, ist für die Beobachtung von Effekten auf den *Extraneous Load* eine Auslastung des *Intrinsic Load* erforderlich (Sweller, 2010, S. 44). Für die Selbsteinschätzung wurden zwar qualifizierte Beispiele dargestellt, es wurde aber nicht kontrolliert, in wie weit sich die Teilnehmenden tatsächlich entsprechend ihrer Kompetenzen einschätzten. Lediglich über die Verteilung der Performanzwerte kann gemutmaßt werden, ob sich die Teilnehmenden entsprechend ihrer Kompetenzen einschätzten. Da die Performanzwerte tendenziell normalverteilt sind (mit wenigen Ausnahmen) ist davon auszugehen, dass der *Intrinsic Load* der Teilnehmenden durchschnittlich nicht zu gering beansprucht war.

Des Weiteren ist zu berücksichtigen, dass die schriftsprachliche Kompetenz der Teilnehmenden unterschiedlich ausgeprägt war. Personen, die gerade erst gelernt haben, Wörter zu lesen, ist die Interaktivität der einzelnen Elemente, die im Arbeitsgedächtnis vorgehalten werden müssen höher als bei Personen, die bereits geübt sind im Lesen. Letztere Personengruppe kann für das Lesen und

Verstehen von Wörtern auf Schemata für ganze Wörter zurückgreifen wohingegen ungeübte Lesende bisher tendenziell Schemata nur für einzelne Buchstabenfolgen entwickelt haben (Sweller, 2010, S. 42 und vgl. Kap. 2.4 zu CLT). Das experimentelle Setting ließ die Kontrolle dieses möglichen Einflussfaktors nicht zu. Denn die Probanden und Probandinnen konnten – unabhängig ihrer tatsächlichen Kompetenzen – das Niveau wählen. Somit ist davon auszugehen, dass der *Intrinsic Load* bei den einzelnen Personen sehr unterschiedlich ausgelastet war und daher keine konstante Größe darstellte, die einen systematischen Effekt zur Folge hätte haben können.

Auch der iterative Seitenaufbau kann ursächlich dafür sein, dass kein Effekt zu identifizieren ist: Neben den visuellen Hinweissignalen kann auch der iterative Seitenaufbau als eine Art Cueing angesehen werden. Durch den schrittweisen Aufbau wurde die Aufmerksamkeit ebenfalls auf die relevanten aktuell zu betrachtenden bzw. zu bearbeitenden Informationen gerichtet, so dass die visuellen Hinweissignale an Effekt verloren haben.

Diskussion über weitere Hinweise auf Auslastungsfaktoren des *Extraneous Load* im Besonderen und des *Cognitive Load* im Allgemeinen

Weitere Hinweise auf Auslastungsfaktoren des *Extraneous Load* und des *Cognitive Load* im Allgemeinen geben die Ergebnisse über a) das Vorwissen sowie b) die subjektiv empfundene Schwierigkeit der Nutzung von otu.lea.

a) Diskussion der Ergebnisse zum Vorwissen

Wie im Theorieteil erläutert (vgl. Kap. 2.4), ist in Untersuchungen zum *Cognitive Load* das Vorwissen einzubeziehen. Das Vorwissen wird in der vorliegenden Studie zum *Cognitive Load* in Form der Computer- und Internetnutzung einbezogen. Auf der Grundlage der Ergebnisse zur Verteilung der Personen mit viel und wenig Computererfahrung auf die Gruppen „Cues an“ und „Cues aus“ ist zwar davon auszugehen, dass das die Computererfahrung in keinem Zusammenhang mit dem Einsatz von Cues steht; jedoch zeigen die Ergebnisse zur Aufgabenbearbeitungszeit, dass die Computererfahrung einen Einfluss auf die Aufgabenbearbeitungszeit hat und somit auf die Auslastung des *Extraneous Load* hindeuten. Dies ist ein weiteres Indiz, in zukünftigen Studien zum *Cognitive Load* der Kritik an der Theorie zu berücksichtigen, in dem das Vorwissen stärker berücksichtigt wird. Fraglich ist beispielsweise, wie das Vorwissen und die Motivation mit den einzelnen Auslastungsbereichen (*Intrinsic Load*, *Extraneous Load* und *Germane Load*) in Wechselwirkung stehen. Des Weiteren werden Studien zu *Cognitive Load* vornehmlich an Institutionen mit Personen durchgeführt, bei denen die kognitiven Fähigkeiten prinzipiell hoch sind (Plass, Kalyuga & Leutner, 2010, S. 81). Die Ergebnisse sind daher nur für bestimmte Bevölkerungsgruppen repräsentativ. Die Usability-Studie mit funktionalen Analphabeten

und Analphabetinnen ist die erste Studie, in der Erhebungen mit dieser Zielgruppe zum Cognitive Load durchgeführt wurden und daher eine Studie, in der Personen mit tendenziell geringen kognitiven Fähigkeiten einbezogen werden. Beispielsweise zeigt die Varianz der Bearbeitungszeit, dass hier Einflussfaktoren eine starke Rolle spielen. Möglicherweise stehen diese in einem Zusammenhang mit dem Cognitive Load. Das führt zu der Annahme, dass gerade Personen mit tendenziell geringen kognitiven Fähigkeiten in Studien zum Cognitive Load einbezogen werden sollten. Im Bereich der Usability-Forschung ist dabei allerdings zu berücksichtigen, dass die ideale Darstellungsform und -art für Personen mit geringen kognitiven Fähigkeiten nicht immer auch ideal für Personen mit tendenziell hohen kognitiven Fähigkeiten darstellt.

b) Diskussion der Ergebnisse zur subjektiv empfundenen Schwierigkeit der Nutzung von otu.lea.

Die Aussagen der Interviews deuten auf die subjektive Wahrnehmung der Beanspruchung des *Extraneous Load* sowie auf weitere Auslastungsfaktoren hin. Die folgende Tabelle zeigt einen Überblick über Effekte, die auf einen Rückschluss auf die Reduktion der Auslastung des *Extraneous Load* bzw. *Cognitive Load* zulassen sowie über die Effekte, die vermutlich eine ansteigende Beanspruchung des *Extraneous Load* bzw. *Cognitive Load* zur Folge haben und daher als Auslastungsfaktoren bezeichnet werden. Die beobachteten Effekte werden jeweils den im Theorieteil erläuterten Annahmen über Auslastungs- bzw. reduzierende Faktoren des Arbeitsgedächtnisses zugeordnet.

Faktoren	Beobachtete Effekte	Theoretische Zuordnung
Reduzierende Faktoren	Instruktionen tragen zur Fokussierung relevanter Inhalte bei (O3)	Extraneous Load
	Iterativer Seitenaufbau (J3)	Extraneous Load
	Möglichkeit des wiederholten Abspielens der Instruktionen und Aufgabeninhalte (J3)	Extraneous Load - Limited Capacity Assumption
	Instruktionen werden permanent angezeigt (E3)	Extraneous Load - Limited Capacity Assumption
	Kombination von Bild und Text (bzw. Ton) erleichtern das inhaltliche Verständnis (AJ3)	Extraneous Load - Dual Channel Assumption
Auslastungs-faktoren	farbliche Gestaltung ist zu wenig kontrastreich (C3)	Extraneous Load
	Informationsdichte ist zu groß. (W3)	Extraneous/Intrinsic Load - Overcomplication
	Aufgaben werden als zu leicht empfunden (V3)	Extraneous/Intrinsic Load - Oversimplification

Tabelle 42: Beobachtung reduzierender und auslastender Faktoren für den Cognitive Load

Ein reduzierender Effekt auf die Auslastung des *Cognitive Load* während der Nutzung von otu.lea kann aufgrund von folgenden, in den Interviews beschriebenen, Faktoren vermutet werden:

Die Instruktionen können dazu beitragen, die relevanten Inhalte zu fokussieren (O3). Damit wird der *Extraneous Load* nicht durch zusätzliche, für die Aufgabenbearbeitung irrelevanten, Auslastungsfaktoren beansprucht. Wie auch bereits in den Überlegungen zu den Cues vermutet, wird der iterative Seitenaufbau als potenziell entlastender Faktor für den *Extraneous Load* genannt (J3). Auch beschreibt diese Person einen Effekt, der mit der *Limited Capacity Assumption* zu erklären ist: Sie erläutert, dass es für sie schwierig ist, die relevanten Informationen im Gedächtnis zu behalten. Daher stellt für sie die Möglichkeit des wiederholten Abspielens der Instruktionen und Aufgabeninhalte eine Entlastung dar (J3). Auch folgende Aussage deutet auf die Annahme hin, dass die Kapazität des Arbeitsgedächtnisses begrenzt ist: Nachdem der Soundbutton aktiviert wurde, bleiben die Textinhalte (hinsichtlich der Instruktionen und des Storyboards) auch während der Aufgabenbearbeitung sichtbar. Somit müssen Informationen nicht im Arbeitsgedächtnis vorgehalten werden, sondern können

immer gelesen bzw. abgespielt werden (E3). Einen Hinweis auf die *Dual Channel Assumption* liefert Person AJ3: Sie beschreibt den positiven Effekt, dass die zusätzliche Darstellung von Informationen über die Bilder das Verständnis erleichtern¹²³.

Als potenzielle Auslastungsfaktoren können folgende Aussagen und Beobachtungen von Probanden und Probandinnen bewertet werden:

Die Person C3 erläutert, dass sie die farbliche Darstellung der Aufgaben als schwierig empfand. Dies scheint bei ihr zu einer starken Beanspruchung des *Extraneous Load* geführt zu haben.

Zudem geben die Interviews weitere Hinweise zur Beanspruchung des Cognitive Load im Allgemeinen. So führt die Informationsdarstellung und -dichte bei der Person W3 zu einer *Overcomplication*: Sie beschreibt, dass sie mit zu vielen Informationen konfrontiert wird und dies zu Verwirrung führt. Zum anderen besteht ein Hinweis auf eine *Oversimplification*: Für die Person V3 ist die Aufgabe so einfach, dass es zu Irritationen kommt.

Über die hier dargestellten Beobachtungen hinaus können des Weiteren die Aussagen zur Heuristik „Perspektivübernahme“ als Hinweise für die Beanspruchung des *Extraneous Load* herangezogen werden (vgl. Kap. 2.4 zu Cognitive Load und Kap. 8.3.7 zu Darstellung der Ergebnisse zum *Extraneous Load*). Von der Mehrheit der Probanden und Probandinnen wurde die Verständlichkeit der Instruktionen bestätigt. Für manche Teilnehmenden waren die Instruktionstexte jedoch zu lang. Hier besteht noch Potenzial, einer starken Beanspruchung des *Extraneous Load* entgegenzuwirken.

In weiteren Untersuchungen wäre neben dem Vorwissen die Motivation systematisch in das Forschungsdesign zu integrieren, was in bisherigen Studien zur CLT und insbesondere zur CTML nur bedingt berücksichtigt wurde (Plass, Kalyuga, u. a., 2010, S. 81). Die Probanden und Probandinnen äußern sich zwar hinsichtlich ihrer Motivation, etwas zu nutzen; es wird jedoch von einigen Probanden und Probandinnen berichtet, dass sich die Motivation im Laufe der Nutzung verändert hat. Die Motivation kann daher nicht als statistischer Einflussfaktor in die Berechnungen und Analysen der qualitativen Daten zum *Extraneous Load* einfließen. Weitere Einflussfaktoren wie Anstrengungsbereitschaft, räumliches Denk- und Vorstellungsvermögen wären in weiteren Studien ebenfalls zu berücksichtigen, um umfassende Aussagen über die Beanspruchung des *Extraneous Load* während der Nutzung der Online-Testumgebung treffen zu

¹²³ Die Beobachtungen sind ein Hinweis auf die Auslastung des *Extraneous Load*, da es mit dem Verständnis und der Umsetzung der Aufgabe vorerst um die Erschließung der Aufgabe (also um die Frage „Wie kann ich die Aufgabe bearbeiten?“) und nicht um die inhaltliche Bearbeitung der Aufgabe (beispielsweise um die Frage „Wird das Wort groß oder klein geschrieben?“) geht.

können. Die Annahme, dass gerade bei Zielgruppen mit diesen Charakteristiken der Effekt von multimedialen Hinweissignalen besonders hoch ist, kann durch diese Studie vorerst nicht bestätigt werden. Um das genauer zu untersuchen wären Studien mit mehr Probanden und Probandinnen und einem speziell auf die Untersuchung des *Extraneous Load* zugeschnittenen Forschungsdesign notwendig.

8.4 Zusammenfassung der Ergebnisse

In der folgenden Abbildung sind sowohl die auf der theoretischen Basis vermuteten als auch die empirisch festgestellten Ergebnisse dargestellt:

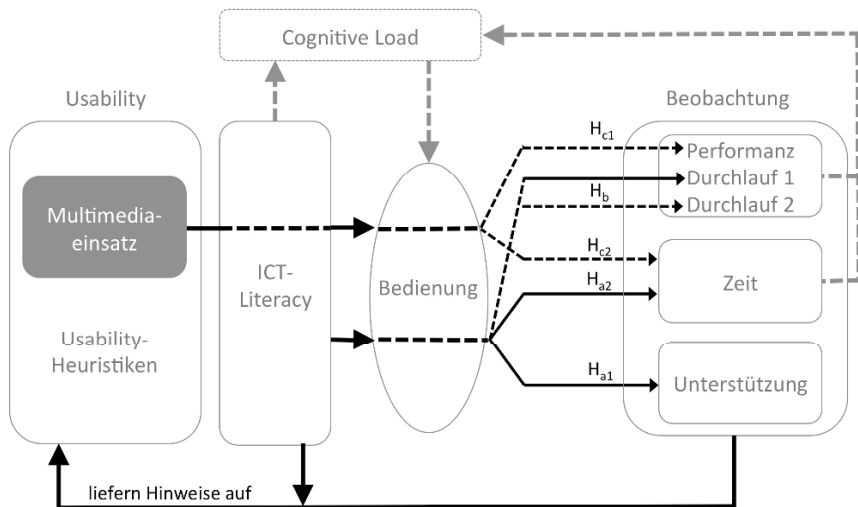


Abbildung 39: Direkte und indirekte empirisch überprüfte Wirkungsbeziehungen

Die Abbildung stellt eine Weiterentwicklung und Konkretisierung der Abbildung 7 (theoretisch vermutete Zusammenhänge) und der Abbildung 28 (theoretische vermutete Zusammenhänge und Forschungshypothesen) dar.

Die Begriffe in den rechteckigen und ovalen Formen bezeichnen die fokussierten und untersuchten Bereiche (Usability-Heuristiken (unter Einsatz von Multimedia), ICT-Literacy, Bedienung von otu.lea, die Performanz, die Zeit, die Unterstützung sowie indirekt die Beanspruchung des Cognitive Load). Die schwarzen Pfeile stellen die geprüften *direkten* Zusammenhänge dar, die grauen Pfeile weisen auf empirisch geprüfte *indirekte* Zusammenhänge hin. Die durchgezogenen schwarzen Linien verdeutlichen einen empirisch festgestellten Zu-

sammenhang (z. B. $H1_{a2}$: Es besteht ein Zusammenhang zwischen der ICT-Literacy – in der Untersuchung abgebildet über die Computererfahrung – und der Bearbeitungszeit) bzw. die Rückschlussfähigkeit von beobachteten Ergebnissen auf die einbezogenen Konzepte (z. B. von den Beobachtungen auf die Beurteilung der Usability). Die perforierten schwarzen Pfeile kennzeichnen empirisch überprüfte aber nicht bestätigte direkte Zusammenhänge. Die perforierten grauen Pfeile deuten geprüfte aber nicht empirisch bestätigte indirekte Zusammenhänge an.

Die Ergebnisse zu den Usability-Heuristiken zeigen, dass die Online-Testumgebung von funktionalen Analphabeten und Analphabetinnen mit wenigen Einschränkungen genutzt werden kann:

- die Heuristiken sind zu einem Großteil erfüllt
- Modifikationsbedarf besteht insbesondere hinsichtlich der Login-Sequenz
- die Teilnehmenden benötigen zu Beginn der Nutzung vermehrte Unterstützung

Weitere Untersuchungen sind hinsichtlich der Wahrnehmungssteuerung und interkultureller Aspekte durchzuführen. Des Weiteren wären Heuristiken einzubeziehen, die in den hier dargestellten Evaluationszyklen aufgrund des Entwicklungsstands nicht berücksichtigt werden konnten (wie beispielsweise die Feedback-Funktion).

Die Computerkompetenz konnte aufgrund einer zu geringen Varianz im Antwortverhalten in weiteren Berechnungen nicht berücksichtigt werden. Die Computererfahrungen sind heterogen und beeinflussen mitunter das Bedienverhalten von otu.lea:

- Sowohl die eingeforderte als auch die nicht eingeforderte Unterstützung werden von der Computererfahrung beeinflusst (Ablehnung der $H0_{a1}$ zugunsten der $H1_{a1}$).
- Die Computererfahrung hat zudem einen Einfluss auf die Bearbeitungszeit der gesamten Testumgebung als auch auf die Aufgabenbearbeitungszeit (Ablehnung der $H0_{a2}$ zugunsten der $H1_{a2}$).
- Auf die Performanz hat die Computererfahrung keinen Einfluss (Bestätigung der $H0_b$) jedoch konnte ein Zusammenhang zwischen der Computererfahrung und der Performanz des ersten Durchlaufs festgestellt werden.

Die experimentellen Bedingungen liefern Hinweise auf Auslastungsfaktoren des *Extraneous Load*:

- Der Einsatz von multimedialer Unterstützung in Form von Cues hat statistisch keinen systematischen Einfluss auf die Performanz (Bestätigung

der $H0_{c1}$). Die Ergebnisse erlauben jedoch Vermutungen hinsichtlich eines unterstützenden Effekts der Cues.

- Der Einsatz multimedialer Cues hat keinen Einfluss auf die Bearbeitungszeit (Bestätigung der $H0_{c2}$).

Somit können keine empirischen Aussagen über den Effekt von Cues auf die Beanspruchung des *Cognitive Load* getroffen werden. Um einen potenziellen Effekt der Cues auf das Bedienverhalten von otu.lea feststellen zu können, wären weitere Studien durchzuführen. Beispielsweise könnte der potenzielle reduzierende Effekt des iterativen Seitenaufbaus auf die Beanspruchung des *Cognitive Load* durch die Einführung einer Kontrollgruppe kontrolliert werden. Bei der Kontrollgruppe würde damit der Seitenaufbau nicht iterativ erfolgen und der Effekt des Einsatzes von Cues könnte ohne diesen zusätzlichen potenziellen Einflussfaktor beobachtet werden.

Die Interviews liefern über die subjektive Einschätzung der Probanden und Probandinnen Hinweise auf Auslastungsfaktoren bzw. reduzierende Faktoren des *Cognitive Load*:

- Die Interviews geben mehrheitlich Hinweise auf Faktoren, die einen reduzierenden Effekt haben (Darstellung und Inhalt der Instruktionen, iterativer Seitenaufbau, Einbindung von Audio und Text).
- Hinweise auf Auslastungsfaktoren betreffen die farbliche Gestaltung, die Informationsdichte sowie die Aufgabenschwierigkeit.

Die Online-Testumgebung wurde vor dem Hintergrund der empirischen Ergebnisse sowie zusätzlicher Modifikationsbedarfe überarbeitet und weiterentwickelt.

8.5 Zyklus 3: Re-Design

Aus dem dritten Evaluationszyklus gehen Modifikationsbedarfe hervor (vgl. Tabelle 36), die soweit möglich in weiteren Entwicklungsschritten realisiert wurden. Zum Zeitpunkt der Erstellung dieser Arbeit konnten als Konsequenz des dritten Evaluationszyklus folgende Modifikationen vorgenommen werden:

1. Einbindung von kontextsensitiven Hilfevideos für jegliche Seitenstrukturen: Für jede Art von Seitenstruktur bzw. Themenbereich sind Hilfevideos implementiert worden.
2. Überarbeitung des Hilfevideos, um das Beenden transparenter zu gestalten: Das Interface hinter dem Hilfevideo wird nicht mehr ausgegraut, sondern komplett ausgeblendet, so dass die Schaltfläche zum Beenden des Hilfevideos prominenter in Erscheinung tritt.
3. Überarbeitung der Instruktionstexte: Die Instruktionstexte wurden partiell überarbeitet. Insbesondere wurden sie bei Möglichkeit gekürzt.

Zudem wurden weitere Modifikationen vorgenommen, die der Entwicklungsplan vorsah oder aus Erfahrungen von Probeeinsätzen mit otu.lea resultierten.

4. Implementation der Feedbackfunktion: Die Feedbackfunktion für Lernende und Kursleitende wurde implementiert:

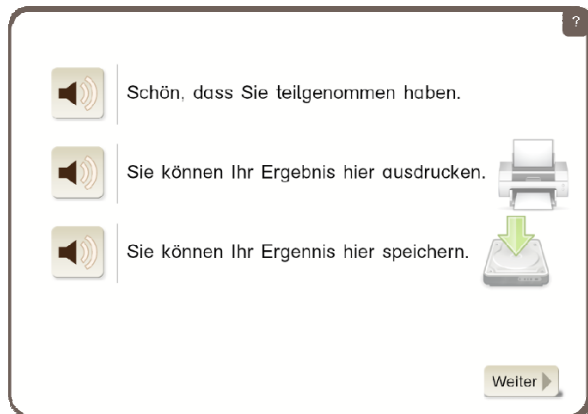


Abbildung 40: Rückmeldung für Teilnehmende

Nach dem Durchlauf können die Teilnehmenden wählen, ob sie sich das Ergebnis anzeigen, ausdrucken oder speichern möchten.

Entscheiden sie sich dafür, das Ergebnis anzeigen zu lassen erhalten sie eine Übersicht über ihre Stärken sowie eine Übersicht über ihren Förderbedarf.

Die Daten werden mit der Statistiksoftware R ausgewertet. Über LATEX werden die PDFs erzeugt.

Schreiben mittel

Teilnahmecode: CYB71

Das kann ich! ☒ Ich beachte auch bei ungebrauchlichen Begriffen die Dopplung des Konsonanten zwischen den Silben. ☒ Ich kann Wörter mit einer Dopplung des Konsonanten zwischen den Silben schreiben. ☒ Ich kann Wörter mit einer Dopplung des Mitlautes im Auslaut richtig schreiben.	Beispiel Pale(t)tjen, Karo(ss)erie Schlüssel, Kasse Schiff, Tipp, Mann

Abbildung 41: Report für Lernende "Das kann ich"

Schreiben mittel

Teilnahmecode: CYB71

Das kann ich bald wenn ich noch ein wenig übe. Ich kann Wörter mit b, k, p oder t schreiben. Ich schreibe groß, was ich sehen oder anfasen kann.	Beispiel Telefon, Paket das Kind, das Flugzeug

Abbildung 42: Report für Lernende „Das kann ich wenn ich noch ein wenig übe“

Für die Kursleitenden gibt es über die Homepage von otu.lea (www.otulea.uni-bremen.de) einen separaten Zugang. Dort können sie einen detaillierten Report über die Lese- und Schreibfähigkeit eines Kursteilnehmenden abrufen¹²⁴.

KANNBESCHREIBUNGEN ERFÜLLT				1	Teilnahme: HB7D6 Datum: 12.05.2014			
LESEN (4)		SCHREIBEN (1)		SPRACHE (2)		RECHNEN (10)		
Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	
1.2.1.1 Kann Wörter mit ansteigender Komplexität (Konsonantenhäufung) recodieren und decodieren	1.2.1 1.2.2	2.2.0.4 Kann Satzschlusszeichen anwenden (Punkt)	2.2.0.1 2.2.0.2 2.2.0.3	3.1.4.1 Kann den possessiven Genitiv in einem kurzen Satz erkennen	3.1.4	4.1.1.1 Kann Mengen im Zehner-Bereich (Z) erfassen	4.1.1	
1.2.1.2 Kann Zeilpläne sinnentnehmend	1.2.1	2.2.0.6 Kann am Anfang des Satzes Großschreibung beachten	2.2.0.2 2.2.0.3	3.2.2.1 Kann verbale Kongruenz im Perfekt in einem mäßig kurzen Satz erkennen	3.2.2	4.1.1.2 Kann Mengen im Hunderter-Bereich (H) erfassen	4.1.1	
KANNBESCHREIBUNGEN TEILWEISE ERFÜLLT				4	Teilnahme: HB7D6 Datum: 12.05.2014			
LESEN (4)		SCHREIBEN (1)		SPRACHE (2)		RECHNEN (10)		
Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	
1.3.4.2 Kann orthographisch komplexere Wörter erfassen	1.3.4 1.3.6	2.4.0.5 Kann a-Laute richtig verwenden (ß, ss)	2.4.0.1 2.4.0.2 2.4.0.3 2.4.0.4	3.1.3.1 Kann nominale Kongruenz im Nominativ in einem kurzen Satz erkennen	3.1.3	4.2.3.1 Kann Einer-Ziffern und Zehner-Zahlen im Kopf zusammenzählen (µ2 K)	4.2.3 4.2.4	
1.3.6.1 Kann einfachen Anleitungen folgen, insbesondere wenn sie Bilder enthalten	1.3.6	2.4.1.1 Kann einen Satz mindestens lautgetreu schreiben	2.4.0.8 2.4.0.9	3.2.1.1 Kann prädikative Kongruenz im Perfekt in einem mäßig kurzen Satz erkennen	3.2.1	4.2.7.1 Kann Hunderter-Zahlen schriftlich zusammenzählen (µ2 S)	4.2.7 4.2.8	
1.4.1.1 Kann einzelne Wörter aus einem Text herausuchen (Item abhängig)	1.4.1 1.4.2 1.4.3 1.4.4 1.4.5 1.4.6	2.5.0.9 Kann Wörter mit á/äu orthographisch richtig schreiben	2.5.0.1 2.5.0.2 2.5.0.3 2.5.0.4	3.3.4.2 Kann orthographisch komplexere Wörter erfassen	3.3.4 3.3.6	4.3.4.1 Kann Einer-Ziffern und Zehner-Zahlen im Kopf voneinander abziehen (µ3 K)	4.3.4 4.3.5	
1.4.1.2 Kann kurzen und einfachen Texten (mit erläuternden Bildern und Illustrationen) 1-2 direkt enthaltene/wörtliche Informationen entnehmen	1.4.1 1.4.2 1.4.3 1.4.4 1.4.5 1.4.6			3.3.6.1 Kann einfachen Anleitungen folgen, insbesondere wenn sie Bilder enthalten	3.3.6	4.3.6.1 Kann Einer-Ziffern und Zehner-Zahlen im Kopf voneinander abziehen (µ3 K)	4.3.6 4.3.7	
KANNBESCHREIBUNGEN NICHT ERFÜLLT				6	12.05.2014			
LESEN (4)		SCHREIBEN (1)		SPRACHE (2)		RECHNEN (10)		
Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	Kannbeschreibung	Aufgabe	
				3.1.1 Kann p-Präsens erkennen	3.1.1 3.1.1.1	4.4.3 Kann Zahlen im Kopf zusammenzählen (µ4 K)	4.4.3 4.4.4	
				3.1.2 Kann ver- in einem	3.1.2 3.1.2.1	4.4.5 Kann Zahlen im Kopf zusammenzählen (µ4 K)	4.4.5 4.4.6	

Abbildung 43: Ausschnitt aus dem Kursleitendenreport von otu.lea

Die Abbildung zeigt einen Ausschnitt eines Kursleitendenreports für den Nutzer/die Nutzerin mit dem anonymen Identifikationscode „HB7D6“. In jedem Report werden die vier Dimensionen Lesen, Schreiben, Sprache und Rechnen abgebildet. Die Zahl in Klammern neben dem Dimensionstitel gibt an, wie oft die Person bisher die Dimension bearbeitet hat. Dem abgebildeten Beispiel zufolge hat die Person bisher viermal die Dimension Lesen, einmal Schreiben, zweimal Sprache und zehnmal die Dimension Rechnen bearbeitet. In dem Report werden jegliche Kannbeschreibungen aufgeführt, die in den von der Person

¹²⁴ Die Abbildung und der Kursleitenden-Report wurden bereits in Koppel und Wolf (2014) veröffentlicht.

bearbeiteten Aufgaben überprüft wurden. Die Kannbeschreibungen werden je nach Performanz drei Kategorien zugeordnet:

Erfüllt: Hier werden die Kannbeschreibungen und Aufgaben gelistet, die erfüllt bzw. richtig bearbeitet wurden.

Teilweise erfüllt: In einem Diagnostikdurchlauf kann eine Kannbeschreibung mehr als einmal geprüft werden; in dieser Kategorie werden die Kannbeschreibungen aufgeführt, die in einer (einigen) Aufgabe(n) erfüllt und in einer (einigen) Aufgabe(n) nicht erfüllt wurde(n).

Nicht erfüllt: Dieser Kategorie werden solche Kannbeschreibungen zugeordnet, deren zugehörige Aufgaben durchgehend falsch beantwortet wurden.

Der Pfeil neben den Nummern der Kannbeschreibungen zeigt an, ob sich die Person im Vergleich zum vorherigen Durchlauf verbessert (Pfeil zeigt nach oben) oder verschlechtert (Pfeil zeigt nach unten) hat. Ist die Performanz gleich geblieben, wird ein waagerechter Pfeil angezeigt. Voraussetzung für die Anzeige des Lernverlaufs ist, dass eine Person mindestens zweimal Aufgaben mit der gleichen Kannbeschreibung bearbeitet hat; ist dies nicht der Fall, wird kein Pfeil angezeigt. In der Spalte rechts neben den Kannbeschreibungen sind die Aufgaben aufgeführt, die bearbeitet wurden. Eine Kannbeschreibung kann in einem Durchlauf und in einer Aufgabe mehr als einmal geprüft worden sein. Das Häkchen neben der Aufgabe gibt an, ob die Kannbeschreibung in der Aufgabe erfüllt wurde. In dem abgebildeten vergrößerten Beispiel für die Dimension Lesen wurde die Kannbeschreibung 1.4.1.1 in sechs Aufgaben (1.4.1, 1.4.2, 1.4.4, 1.4.5, 1.4.6) geprüft; in der Aufgabe 1.4.9 wurde die Kannbeschreibung nicht erfüllt, in den weiteren fünf Aufgaben wurde die Kannbeschreibung erfüllt. Im Vergleich zum vorherigen Durchlauf hat sich die Person verbessert.

5. Implementation der Option, jederzeit auszuwerten und den Test zu beenden oder zur Aufgabenbearbeitung zurückzukehren: Sobald eine Aufgabe bearbeitet wurde besteht die Möglichkeit, in der oberen Leiste des Bildschirms „Zur Dimensions- und Niveauauswahl“ sowie „Auswertung“ auszuwählen.

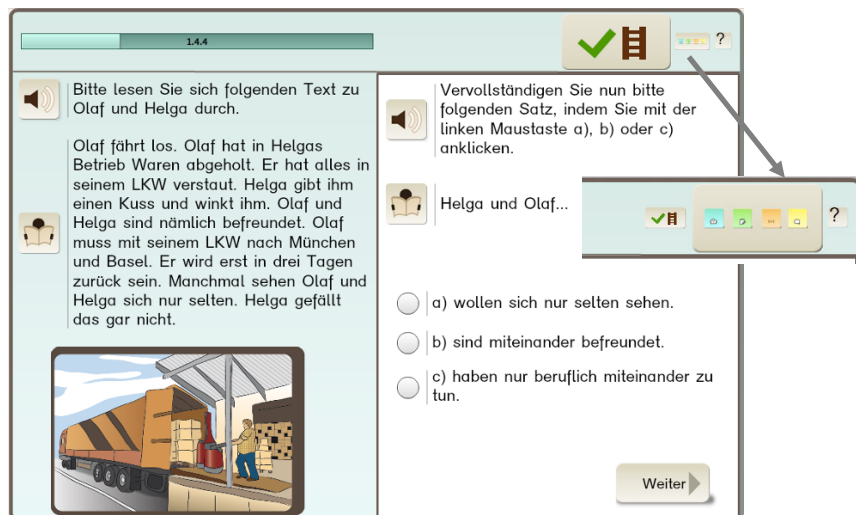


Abbildung 44: Möglichkeiten zur sofortigen Unterbrechung/Auswertung

Mit diesen Optionen können Teilnehmende vor dem Testende die Testumgebung verlassen, sich die Aufgaben auswerten lassen und ggf. zu einem späteren Zeitpunkt die Bearbeitung fortsetzen. Des Weiteren wird ein Fortschrittsbalken angezeigt, der das Verhältnis zwischen bereits bearbeiteten und noch zu lösenden Aufgaben anzeigt.

Insgesamt dauerte der Entwicklungsprozess von otu.lea von ersten konzeptionellen Entwürfen bis hin zur gegenwärtig bestehenden Vollversion fünf Jahre. Die Online-Testumgebung ist unter www.otulea.uni-bremen.de abrufbar. Um otu.lea weiterhin kontinuierlich zu verbessern und an die Herausforderungen der Alphabetisierungspraxis anzupassen, sind weitere Evaluationsstudien geplant. Die Schwerpunkte sollen dabei einerseits auf der Prozessangemessenheit liegen, um die Online-Testumgebung im institutionellen Kontext als effizientes Instrument in Kursen zu etablieren. Andererseits sollen weitere Untersuchungen zum anonymen Einsatz durchgeführt werden, um auch dem Großteil der funktionalen Analphabeten und Analphabetinnen außerhalb von Alphabetisierungskursen die Nutzung lernmotivationssteigernd zu ermöglichen.

8.6 Zusammenfassende Diskussion und Schlussfolgerung

Im theoretischen Teil dieser Arbeit wurde dargelegt, dass in Deutschland ein enormer Alphabetisierungsbedarf besteht, jedoch geeignete Diagnose- und Förderinstrumente für die Praxis fehlen. Mit der Entwicklung der Online-Diagnostik

otu.lea wurde diesem Bedarf begegnet. Herausfordernd war bei der Entwicklung, dass bei der Zielgruppe der funktionalen Analphabeten und Analphabetinnen nicht nur von einer wenig ausgeprägten Literalität, sondern auch einer niedrigen ICT-Literacy ausgegangen werden musste. Zudem gibt es nur wenige Erkenntnisse über eine gerbrauchstaugliche Gestaltung von Interfaces für erwachsene wenig literalisierte Personen. Es galt, eine Online-Testumgebung zu entwickeln, bei der ein möglicher Einfluss der ICT-Literacy auf das Testergebnis minimiert bzw. verhindert und die Auslastung des *Extraneous Load* möglichst gering gehalten werden. Realisiert werden sollte das Ziel durch Berücksichtigung aktueller Erkenntnisse der jeweiligen Forschungsgebiete: Computerbasierte Kompetenzdiagnostik, CLT bzw. CTML und Usability. Mit den Usability-Heuristiken konnten Evaluationskriterien entwickelt werden, auf dessen Grundlage das Instrument entwickelt und überprüft werden konnte. Neben laufenden Evaluationen mit Experten und Expertinnen wurden drei aufeinander aufbauende Evaluationszyklen durchgeführt: Ein Paper Prototyping-Experiment in einem frühen Entwicklungsstadium im Jahr 2010, eine Usability-Studie mit einem Prototyp im Jahr 2012 sowie eine abschließende umfangreiche Usability-Studie im Jahr 2013.

Der DBR-Ansatz stellte aufgrund seiner flexiblen Eigenschaften einen idealen Rahmen für das Vorhaben, eine Online-Testumgebung für funktionale Analphabeten und Analphabetinnen zu entwickeln und diese systematisch iterativ zu evaluieren, dar. Dabei wurden die jeweiligen Forschungsperspektiven – sowohl kognitionspsychologische als auch (kompetenz-) diagnostische – Perspektiven einbezogen und mit Erkenntnissen der Usability-Forschung verknüpft. Die übergeordneten Ziele des Forschungsvorhabens waren (A) einen theoretischen Output mit kontextualisierten Theorien sowie (B) einen praktischen Output – die Online-Testumgebung für funktionale Analphabeten und Analphabetinnen – zu generieren. Das Ziel des theoretischen Outputs ist wiederum in drei Teilbereiche zu differenzieren (vgl. Abbildung 15 im Kap. 5.2 zu Zielen des Forschungsvorhabens):

A1: Entwicklung einer bereichsspezifischen Theorie über computerbasierte Diagnostik im Grundbildungsbereich,

A2: Entwicklung eines *Design-Frameworks* für die Gestaltung einer Online-Testumgebung im Grundbildungsbereich,

A3: Entwicklung einer *Design Methodologie* für den Ablauf von Gestaltungsprozessen bei der Entwicklung von Software für funktionale Analphabeten und Analphabetinnen.

In Anbetracht der Ergebnisse können diese Ziele bedingt erreicht werden: Die Forschungsbemühungen und die gewonnenen Ergebnisse erlauben aufgrund ihrer Diversität keine Ableitung diskursfähiger *Theorien*. Jedoch ist es möglich,

Handlungsempfehlungen zu formulieren. Die Ziele werden daher nicht in ihrer ursprünglichen Formulierung aus Kap. 3.1 und Kap. 5.2 übernommen, sondern entsprechend der eben angestellten Überlegungen umbenannt. So werden im Folgenden die Ergebnisse dahingehend reflektiert, bereichsspezifische *Handlungsempfehlungen* für die Entwicklung einer computerbasierten Diagnostik, *Empfehlungen* für ein *Design Framework* sowie *Empfehlungen* für eine *Design Methodology* zu formulieren. Den Handlungsempfehlungen werden jeweils Vorbemerkungen, in denen die Ergebnisse zusammengefasst und reflektiert werden, vorangestellt.

Ziel A1: Entwicklung einer bereichsspezifischen Handlungsempfehlung über computerbasierte Diagnostik im Grundbildungsbereich

Vorbemerkung: Die Herausforderungen der Entwicklung einer computerbasierter Diagnostik im Grundbildungsbereich bestanden primär in einer tendenziell niedrigen ICT-Literacy sowie der niedrigen literalen und mathematischen Kompetenzen. Ziel bei der Entwicklung der computerbasierten Diagnostik war, dass die ICT-Literacy und die Literalität keinen Einfluss auf die Performanz ausüben. Bei der Entwicklung wurde daher auf eine ausgeprägte Gebrauchstauglichkeit und insbesondere leichte Erschließbarkeit (sowohl der Testumgebung als auch des Inhalts der Testumgebung) durch multimediale Unterstützung geachtet. Die Ergebnisse hinsichtlich eines möglichen Zusammenhangs zwischen der Computererfahrung und der Bedienbarkeit erlauben keine eindeutige Antwort. Die Bedienbarkeit wurde über die eingeforderte Unterstützung (1) und die Aufgabenbearbeitungszeit (2) operationalisiert:

- 1) Bezüglich der eingeforderten Unterstützung besteht ein Zusammenhang zwischen den Computererfahrungen und der eingeforderten Unterstützung: Je niedriger die Computererfahrung, desto öfter wird der Proband/die Probandin unterstützt.
- 2) Für die Aufgabenbearbeitungszeit wurde festgestellt, dass sich für Formate auf höherem Niveau und tendenziell für offene Formate ein Zusammenhang zwischen der Computererfahrung und der Bearbeitungszeit abzeichnet. Ein Zusammenhang zwischen der Computererfahrung und der formatunabhängigen Bearbeitungszeit kann aber nicht bestätigt werden. So ist davon auszugehen, dass im Kontext der durchgeführten Studien, die Computererfahrung tendenziell keinen Einfluss auf die Bearbeitungszeit hat.

Auch die Frage, ob ein Zusammenhang zwischen der Computererfahrung und der Performanz besteht, konnte nicht eindeutig beantwortet werden: Es scheint zwar ein Zusammenhang zwischen der Computererfahrung und der Performanz zu bestehen. Allerdings ergeben die formatbezogenen Performanzwerte für den

separaten zweiten Durchlauf keine signifikanten Werte, was in der geringeren Fallzahl begründet sein kann. Es wird aber deutlich, dass für die Items auf den schwierigsten Niveaus (Lesen schwer und Schreiben mittel) tendenziell ein niedrigerer Performanzwert erzielt wurde als auf den einfacheren Niveaus, nur das halboffene Format im Bereich Schreiben auf dem einfachen Niveau ist dieser Tendenz nicht zuzuordnen.

Zusammenhänge zwischen Cues und der Performanz sowie zwischen Cues und der Bearbeitungszeit können nicht festgestellt werden. Auch gibt es wenig Hinweise auf der Grundlage der Interviewaussagen auf die Beanspruchung des *Extraneous Load*. Vereinzelt äußerten sich Teilnehmende über die subjektiv wahrgenommene Schwierigkeit des Programms, was als Hinweis auf den *Extraneous Load* gedeutet werden kann, doch kann keine generelle Tendenz hin zu einer starken oder niedrigen Beanspruchung festgestellt werden. Dass die Probanden und Probandinnen mehrheitlich keine Schwierigkeiten erwähnen, kann als Tendenz hin zu einer niedrigen Beanspruchung des *Extraneous Load* gedeutet werden.

Bereichsspezifische Handlungsempfehlung über eine computerbasierte Diagnostik im Grundbildungsbereich:

Bei der Entwicklung einer computerbasierten Diagnostik im Grundbildungsbereich Erwachsener ist von heterogenen Computererfahrungen auszugehen. Die Bereitschaft, mit dem Computer zu arbeiten, ist tendenziell vorhanden (s. Ergebnisse zu Computererfahrung). Die Einflussnahme durch Computerkompetenz und/oder -erfahrungen auf das Testergebnis sind zu vermeiden. Testangst und Schamhaftigkeit sind bei der Entwicklung zu berücksichtigen. Hemmschwellen und Nervosität sind beispielsweise durch Übungssequenzen vor der performanzrelevanten Testung abzubauen.

Ziel A2: Entwicklung eines **Design Framework** für die Gestaltung einer Online-Testumgebung im Grundbildungsbereich

Vorbemerkung: Die Ergebnisse zu den Usability-Heuristiken zeigen, dass in der Online-Diagnostik etwa sechs der zehn überprüften Heuristiken (Erlernbarkeit, Aufgabenangemessenheit, Erwartungskonformität, Steuerbarkeit, Prozessangemessenheit, Joy of Use) erfüllt und zwei (Selbstbeschreibungsfähigkeit, Perspektivübernahme) teilweise erfüllt sind. Zwei Heuristiken (Wahrnehmungssteuerung, Interkulturelle Aspekte) sind aufgrund mangelnder Aussagekraft der Ergebnisse nicht beurteilbar. Die Selbstbeschreibungsfähigkeit der Login-Sequenz ist zu überarbeiten, die Instruktionstexte wurden bereits überarbeitet, um die Perspektivübernahme zu verbessern. Des Weiteren ist aus der Evaluation abzuleiten, dass die überprüften Heuristiken (auch) für funktionale Analphabeten

und Analphabetinnen gelten. Wurden Heuristiken nur teilweise erfüllt, hatte das zur Folge, dass Irritationen, falsche Eingaben (z. B. bei der Login-Sequenz), gesteigerter Unterstützungsbedarf und/oder Bearbeitungsverzögerungen auftraten (s. z. B. Ergebnis zur Heuristik Selbstbeschreibungsfähigkeit). Diese Ergebnisse deuten zudem darauf hin, dass bezüglich bestimmter Heuristiken Konkretisierungen vorzunehmen sind. Im Folgenden wird auf der Grundlage dieser Ergebnisse ein *Design Framework* in Form von Usability-Heuristiken abgeleitet.

Handlungsempfehlung für ein Design Framework - Usability-Heuristiken für funktionale Analphabeten und Analphabetinnen:

Grundsätzliche Gestaltungshinweise (vgl. Usability-Heuristiken Kap. 2.6.4) für die gebrauchstaugliche Gestaltung von Interface sind zu berücksichtigen. Allerdings bedürfen die Heuristiken zum Teil einer Konkretisierung bzw. Ergänzung. Die Ergänzungen zu den bereits bestehenden Heuristiken werden im Folgenden aufgeführt:

Erlernbarkeit: Um eine Testumgebung möglichst erlernbar zu gestalten, sind verstärkt zu Beginn unterstützende Hinweise für die Bearbeitung der Umgebung einzubinden.

Feedback¹²⁵: Neben der Hilfefunktionen können weitere Hinweise z. B. unmittelbare Feedbackfunktionen in Form von Pop Up-Fenstern bei falschen Eingaben (hinsichtlich der nicht inhaltsbezogenen Bearbeitung) oder ab einer bestimmten Zeitspanne ohne Interaktion sein.

Selbstbeschreibungsfähigkeit: Für eine ausgeprägte Selbstbeschreibungsfähigkeit ist der Seitenaufbau iterativ zu gestalten und es sind Hinweise für die Bearbeitung der Umgebung einzubinden. Die Unterstützungsfunktionen sind kontextsensitiv zu gestalten, so dass sie an die jeweilige Umgebung, in der sich ein Nutzer/eine Nutzerin befindet, angepasst sind.

Steuerbarkeit: Aktivierbare Schaltflächen sind in ihrer Funktion zu erläutern. Insbesondere bei der Einbindung von Videos und Animationen ist darauf zu achten, Start- und Ende-Schaltflächen deutlich darzustellen und prominent zu platzieren. Jegliche Instruktionen und Unterstützungen müssen beliebig oft abrufbar sein.

Perspektivübernahme: Instruktionstexte sind kurz und verständlich zu formulieren.

Als zusätzliche Heuristik neben bereits existierenden Heuristiken ist für die Gestaltung von computerbasierten Testumgebungen eine weitere Heuristik einzuführen:

¹²⁵ Auch wenn die Heuristik Feedback nicht Bestandteil der Evaluation war, lassen die Beobachtungen Schlussfolgerungen hinsichtlich dieser Heuristik zu.

Übungssequenz: Bei der Gestaltung von computerbasierten Testumgebungen sind Übungssequenzen vorzuschalten, durch die sich Teilnehmende mit grundsätzlichen Funktionen vertraut machen können.

Die Empfehlungen richten sich in erster Linie an die Entwicklung computerbasierter Diagnoseinstrumente für den Grundbildungsbereich Erwachsener. Darüber hinaus ist zu prüfen, ob die Heuristiken grundsätzlich auf die Gestaltung von Testumgebungen für Personen mit geringen Computererfahrungen und -kompetenzen anwendbar sind.

Diese Handlungsempfehlungen sind, wie jegliche Anwendungen von Heuristiken (vgl. Kap. 2.6.4.3 zu Grenzen der Anwendbarkeit von Heuristiken), kontextabhängig zu betrachten. Da aber diese Heuristiken explizit auf eine bestimmte Zielgruppe bezogen sind, trifft der Kritikpunkt an Heuristiken – einer zu starken Generalisierung – hier nur bedingt zu. Auch die Gefahr einer Missinterpretation und als Folge dessen die Übertragung auf einen nicht adäquaten Nutzungskontext ist nur gering, da die Überprüfung und Entwicklung der Heuristiken transparent und intersubjektiv nachvollziehbar dargestellt wurde.

Ziel A3: Entwicklung von *Design Methodologies* – Ablauf von Gestaltungsprozessen bei der Entwicklung von Software für funktionale Analphabeten und Analphabetinnen

Vorbemerkung: Für die Entwicklung der Online-Testumgebung wurden verschiedene Evaluationsmethoden herangezogen. Die Gestaltung der Evaluationszyklen sowie der Instrumente und der Einsatz der Erhebungsmethoden wurde vor dem Hintergrund theoretischer Annahmen getroffen. Wie sich in der Praxis herausstellte, mussten Erhebungsinstrumente angepasst werden (z. B. der Interviewleitfaden zwischen dem zweiten und dem dritten Evaluationszyklus). Zielgruppenspezifische Charakteristika führten zudem zum Teil zu einer begrenzten Aussagekraft der Ergebnisse. Aus diesen Erfahrungen können Annahmen über erfolgsbedingende Faktoren getroffen und als Handlungsempfehlungen für Design Methodologies formuliert werden.

Handlungsempfehlung für eine Design Methodology für den Ablauf von Gestaltungsprozessen bei der Entwicklung von Software für funktionale Analphabeten und Analphabetinnen:

Rückblickend ist die Berücksichtigung der DBR-Merkmale eine Grundlage für das Gelingen des Evaluationsprozesses. Zwei Bereiche sollen dabei im Folgenden herausgestellt und konkretisiert werden, da sie eine besondere Relevanz in dem Forschungsprozess besitzen.

- 1) Potenzielle Nutzer und Nutzerinnen sind von Beginn an in den Entwicklungsprozess einzubeziehen. So war es einerseits bedeutsam, funk-

tionale Analphabeten und Analphabetinnen am Evaluationsprozess zu beteiligen, um Hinweise für die gebrauchstaugliche Gestaltung zu erhalten. Zudem waren aber auch Kursleitende wichtige Entwicklungspartner und -partnerinnen, da diese einerseits ihre Erfahrungen mit der Zielgruppe einbringen und andererseits die Bedarfe und Herausforderungen in der Alphabetisierung identifizieren und in Handlungsempfehlungen hinsichtlich der Gestaltung von otu.lea transferieren können. Diese sind wiederum Bestandteil der Reflexion der Erhebungsinstrumente und es wird an dieser Stelle auf das Kap. 9.2 verwiesen.

- 2) Die Wahl der Erhebungsinstrumente und -methoden sowie dessen Einsatz sind an die Zielgruppe anzupassen. In dem Kap. 8.2 wird der Einsatz der Erhebungsinstrumente ausführlich reflektiert. Im Folgenden wird eine Zusammenfassung der Handlungsempfehlungen für den Einsatz der Erhebungsinstrumente dargestellt (an dieser Stelle sei zudem auf das Kap. 9.2 verwiesen, in dem die Handlungsempfehlungen vor dem Hintergrund einer ausführlichen Reflexion der Erhebungsinstrumente abgeleitet werden):

- Bei der Durchführung von Paper Prototyping-Experimenten sind zu Beginn des Experiments eine Übungssequenz einzubinden, in der Vertrauen zwischen Forscher/Forscherin und Proband/Probandin aufgebaut und die Spielregeln verinnerlicht werden können.
- Für die Durchführung von Usability-Tests sind den Betreuern und Betreuerinnen verbindliche Anweisungen zu geben, z.B. eine bestimmte Zeitspanne vorzugeben, bevor sie eingreifen, um die Handlungen der Probanden und Probandinnen nicht zu beeinflussen und somit vergleichbare Ergebnisse zu erzielen.
- Interviewfragen sind aufgrund des Risikos der Sozialen Erwünschtheit offen zu stellen und sollten zudem an dem Erfahrungswissen der Teilnehmenden anknüpfen. Für eine valide Datenbasis kann ergänzend eine kommunikative Validierung hilfreich sein.
- Befragungen mit Fragebögen, in denen die Antwortmöglichkeiten nur in ihren extremen Ausprägungen vorgegeben sind, sind entweder zu vermeiden oder es sind qualifizierende Beispiele bzw. Operationalisierungen der Abstufungen vorzunehmen.

Ziel B: *Praktischer Output* – Online-Testumgebung für funktionale Analphabeten und Analphabetinnen

Vorbemerkung: Das Ziel wurde durch die Entwicklung von otu.lea erreicht. Es besteht im Einzelnen Modifikationsbedarf, jedoch ist die Testumge-

bung voll funktionsfähig und weist eine ausgeprägte Gebrauchstauglichkeit auf. Sie ist somit sowohl anonym als auch in Alphabetisierungskursen einsetzbar.

Praktischer Output: Online-Testumgebung für funktionale Analphabeten und Analphabetinnen:

Für die genaue Beschreibung der Online-Testumgebung otu.lea s. Kap. 4 im theoretischen Teil zur Entwicklung der Online-Testumgebung und Kap. 8.5 zu Re-Design des dritten Zyklus.

Zusammenfassend ist die Mehrheit der überprüften Heuristiken erfüllt, empirisch hat die Computererfahrung zwar einen Einfluss auf die Unterstützung, jedoch weder auf die Aufgabenbearbeitungszeit noch auf die Performanz. Hinweise auf eine starke Beanspruchung des *Extraneous Load* konnten nicht festgestellt werden. Schlussfolgernd kann die Usability von otu.lea als ausgeprägt aber in wenigen Aspekten als modifikationsbedürftig beurteilt werden.

Phase IV – Reflexion

9 Reflexion der Erhebungsmethoden und -instrumente

Im Folgenden werden der Forschungsrahmen DBR sowie die Erhebungsmethoden und -instrumente ausführlich reflektiert. Im Anschluss an die Reflexion der Erhebungsinstrumente werden jeweils Handlungsempfehlungen für den Einsatz der entsprechenden Methoden formuliert.

9.1 Reflexion des Forschungsrahmens DBR

Für die Entwicklung und Evaluation der Online-Testumgebung otu.lea stellte der DBR-Ansatz einen tragfähigen Forschungsrahmen dar. Durch mehrere Design-Zyklen konnten unterschiedliche Bereiche fokussiert und verschiedene Evaluationsmethoden eingesetzt werden. Die iterative Vorgehensweise ermöglichte die Berücksichtigung und Überprüfung von Aspekten, die ggf. im Vorfeld nicht vorgesehen waren. Dadurch, dass der DBR-Ansatz nicht durch festgelegte Methoden definiert ist, konnten je nach Kontext und fokussierter Fragestellung die Erhebungsmethoden und -instrumente ausgewählt werden. Dies war insbesondere bezüglich der Zielgruppe von Relevanz: Zu zielgruppenspezifischen Charakteristika im Forschungskontext konnten im Vorfeld aufgrund mangelnder Erfahrungen nur Annahmen getroffen werden. Die Erfahrungen zeigten, dass zielgruppenspezifische Charakteristika (Soziale Erwünschtheit und Zustimmungstendenz) insbesondere in Interviewsituationen stärker zu berücksichtigen sind. Die Erhebungsmethoden, wie beispielsweise der Interview-Leitfaden, konnten daher nach dem Einsatz im zweiten Zyklus modifiziert und an die Herausforderungen der Interviewsituation bedingt durch Charakteristika der Zielgruppe angepasst werden. Somit konnte eines der beiden im Kontext von DBR formulierten Ziele – die Generierung eines praktischen Outputs, der eine Lösung für ein Problem in der Praxis liefert – erreicht werden.

Das weitere Ziel des DBR, die Generierung eines theoretischen Outputs, konnte bedingt erreicht werden. Die Datenlage bietet nur eingeschränkt die Möglichkeit der Generalisierung. Folglich ist auch die Entwicklung von bereichsspezifischen Theorien nicht vollständig möglich. Im Kontext dieser Studie war das Ziel des theoretischen Outputs, die Entwicklung und Überprüfung von Usability-Heuristiken. Die Überprüfung der Heuristiken war stark domänen- und kontextspezifisch, denn die Heuristiken wurden mit Kursteilnehmenden (und nicht auch mit funktionalen Analphabeten und Analphabetinnen, die nicht an Alphabetisierungskursen teilnehmen, vgl. Reflexion der Stichprobenauswahl im folgenden Kap. 9.3) und mit nur einem Instrument – otu.lea – überprüft. Dennoch können

die Heuristiken als Diskussions- und Ausgangsgrundlage für weitere Forschungsaktivitäten dienen. Weitere grundsätzliche Kritikpunkte am DBR-Ansatz (Involviertheit der Forschende und Praktizierende sowie Kausalattribution, vgl. Kap. 3.1.6 zu DBR) wurden durch das Forschungsdesign entkräftet: Durch die Verwendung etablierter Erhebungs- und Auswertungsmethoden sowie die Einhaltung der Gütekriterien konnte die Involviertheit der Forschende reflektiert und in der Datenauswertung und -interpretation berücksichtigt werden. Des Weiteren gaben die Videos Aufschluss über die mögliche Einflussnahme durch die Forscher und Forscherinnen bzw. Betreuer und Betreuerinnen. Zudem konnten durch das teilweise experimentelle und hypothesenüberprüfende Design kausale Rückschlüsse gezogen werden. Das Risiko eines narrativen Charakters der Ergebnisse wurde durch die Aufzeichnung der Studie mit Tonband- und Videogeräten, der laufenden Kommunikation und Präsentation in der Fachwelt (durch Publikationen und Tagungspräsentationen) und der Überprüfung der InterCoder-Reliabilität hinsichtlich der Interviewauswertung minimiert. Auch wenn die systematische Dokumentation und die Durchführung mehrerer Evaluationszyklen einen hohen Aufwand im Vergleich zu summativen Evaluationen bedeuten können¹²⁶, ist der Mehrwert eines iterativen Vorgehens unter Anwendung des DBR-Ansatzes zu betonen: Durch die Anwendung des DBR-Ansatzes wurde die Entwicklung des Online-Diagnoseinstruments vorerst auf theoretischer Basis unter Berücksichtigung aktueller Forschungsergebnisse umgesetzt. Der Design-Prozess wurde somit systematisch mit der theoretischen Basis verknüpft. Die Evaluationszyklen fanden unter Einsatz von Forschungsmethoden statt, die sowohl entsprechend des Entwicklungsstadiums als auch der Zielgruppe gewählt und ggf. modifiziert wurden. Damit konnte sich nicht nur iterativ an ein optimales Entwicklungsergebnis herangetastet, sondern auch die Erhebungsmethoden entsprechend der Rahmenbedingung optimiert werden. Des Weiteren ermöglicht der DBR-Ansatz die Verknüpfung von einem hypothesengeleiteten mit einem explorativen Vorgehen und damit die systematische Beobachtung und Berücksichtigung von möglichen Einflussfaktoren, was mit einem rein explorativen Vorgehen nur bedingt möglich ist. Schlussfolgernd bot der DBR-Ansatz einen Forschungsrahmen, um vor dem Hintergrund aktueller theoretischer Erkenntnisse (zu pädagogischer computerbasierter Diagnostik, zum funktionalen Analphabetismus, zur *CLT* sowie Usability-Forschung) Probleme in der Praxis (Herausforderungen in der Alphabetisierung) in Zusammenarbeit mit Forschenden und Praktikern/Praktikerinnen zu lösen, den Prozess systematisch zu reflektieren und

¹²⁶ Die einzelnen Zyklen werden erst auf der Grundlage vorheriger Zyklen konzipiert, die Erhebungsmethoden müssen an den Entwicklungsstand angepasst werden, ggf. sind neue Erhebungsmethoden einzusetzen. Der Einsatz mehrerer iterativer Evaluationszyklen bedeutet folglich einen konzeptionellen und organisatorischen Aufwand, der personelle und finanzielle Ressourcen bindet.

somit sowohl einen Beitrag auf praktischer als auch auf theoretischer Ebene zu leisten: Die Entwicklung einer gebrauchstauglichen Online-Diagnostik sowie die Überprüfung bzw. Neuformulierung von Usability-Heuristiken und die Formulierung von Handlungsempfehlung zum Einsatz von Erhebungsmethoden mit funktionalen Analphabeten/Analphabetinnen.

9.2 Kritische Reflexion der Erhebungsmethoden und –instrumente¹²⁷

In den drei Evaluationszyklen wurde der formal-analytische Messansatz (mittels einer Zielgruppenanalyse und einer Befragung zur Computererfahrung und -kompetenz), der interaktionszentrierte Messansatz (anhand von Beobachtungsprotokollen und Data-Logging) und der benutzer-/benutzerinnenzentrierte Messansatz (mit fokussierten Interviews sowie einer Befragung mit dem UEQ zur subjektiven Bewertung der Gebrauchstauglichkeit) verwendet. Im Folgenden werden die Erhebungsmethoden und -instrumente sowie die Stichprobenauswahl einer kritischen Reflexion unterzogen.

Reflexion der Erhebungsmethoden und -instrumente

Die folgende Tabelle zeigt eine Übersicht der durchgeführten Untersuchungen, systematisiert nach den Methoden und Messansätzen:

Methode	Messansatz	Jahr	N
(a) Paper Prototyping	interaktionszentriert	2010	5
(b) Bearbeitung von Aufgabensamples (Usability-Testing)	interaktionszentriert	2012 und 2013	11 und 50
(c) Fokussierte Interviews	benutzer-/benutzerinnenzentriert	2012 und 2013	11 und 50
(d) Fragebogen zur Computererfahrung und -kompetenz	formal-analytisch	2013	50
(e) User Experience Questionnaire	benutzer-/benutzerinnenzentriert	2013	50

Tabelle 43: Übersicht der eingesetzten Erhebungsmethoden und Messansätze

Die Usability-Studien mit otu.lea bestanden aus drei Evaluationszyklen. Nachfolgend werden die Methoden aus den beiden Vorstudien sowie der Hauptuntersuchung zunächst noch einmal kurz erläutert, um anschließend Handlungsempfehlungen für deren Einsatz zu formulieren.

¹²⁷ Die Reflexion der Erhebungsinstrumente wurde bereits in leicht veränderter Form in Koppel, Küster und Wolf (2013) veröffentlicht.

Das Paper-Prototyping (a) bestand aus der Bearbeitung von vier Aufgaben. Der Usability-Test (b) fand in einem Raum mit Computerarbeitsplätzen statt, an denen die Teilnehmenden parallel arbeiteten. Jeder Monitor wurde mit einer Kamera aufgezeichnet und es fand ein Data Logging (Performanz und Bearbeitungszeit) statt. Anschließend wurden leitfadengestützte fokussierte Interviews durchgeführt (c). Im Jahr 2013 wurden die Teilnehmenden im Anschluss an die Aufgabenbearbeitung zu ihrer Computererfahrung und -kompetenz sowie (d) mit dem UEQ (e) befragt. Der Fragebogen zur Computerkompetenz bestand aus Items zur erforderlichen Computerkompetenz. Die Abfrage zur Computererfahrung bezog sich auf das Nutzungsverhalten; die Fragebögen wurden den Teilnehmenden vorgelesen.

Potenzielle Störvariablen waren die Instruktionen der Betreuer und Betreuerinnen, die Formulierung der Interviewfragen sowie die Computererfahrung/-kompetenz. Um den Einfluss der Störvariablen zu reduzieren, wurden die Betreuer/Betreuerinnen und Interviewer/Interviewerinnen im Vorfeld detailliert instruiert und von Experten und Expertinnen hinsichtlich der Interviewdurchführung geschult. Kontrolliert werden konnten diese Störvariablen durch die Aufzeichnung der Usability-Tests mit Videokameras sowie der Aufzeichnung der Interviews mit Tonbandgeräten. Die Computererfahrung sowie -kompetenz wurden mittels einer Befragung erhoben und in die Auswertung der Ergebnisse einbezogen. Weitere typische Störvariablen in Untersuchungssettings wie beispielsweise Temperatur, Uhrzeit und Lärm, konnten aufgrund des gesetzten Settings (der Alphabetisierungskurse) im Vorfeld nicht beeinflusst werden. Da die Probanden und Probandinnen Kopfhörer trugen ist allerdings davon auszugehen, dass der Faktor Lärm keine ergebnisbeeinflussende Störvariable darstellt. Auch Raumtemperatur und Zeitpunkt der Untersuchung schienen die Aufmerksamkeit und/oder Konzentration nicht zu beeinträchtigen.

Im Folgenden werden die Erfahrungen mit den verschiedenen Methoden des Usability-Testing mit funktionalen Analphabeten und Analphabetinnen reflektiert und mit entsprechenden Empfehlungen versehen.

(a) Paper Prototyping

Es ergab sich die Schwierigkeit des Transfers zwischen dem papierbasierten Interface und der gedanklichen Übertragung des Effekts der eigenen Handlung auf einen Bildschirm. Zunehmend wurden jedoch die „Spielregeln“ des Paper Prototypings verinnerlicht. Da es sich beim Paper Prototyping um eine qualitative Methode handelt, die zum Ziel die Entwicklung eines Produkts und nicht die Generierung allgemeingültiger Erkenntnisse zum Ziel hat, sind die möglichen Einflussfaktoren durch Nervosität, mangelndes Verständnis oder geringe Transfermöglichkeit tolerierbar. Dennoch wäre der Testsituation (Abbau von Hemmschwellen und Nervosität) und der Ergebnisqualität zuträglich, *zu Beginn des*

Experiments eine Übungssequenz einzubinden, in der Vertrauen zwischen Forscher/Forscherin und Proband/Probandin aufgebaut und die Spielregeln verinnerlicht werden können.

(b) Bearbeitung von Aufgabensamples (Usability-Testing)

Auch wenn sich die Betreuungspersonen im Hintergrund aufhielten, löste die Situation bei einigen Teilnehmenden Stress aus. Das äußerte sich in zitternden Händen, leichten Schweißausbrüchen oder verbalen Äußerungen im Interview. Abhilfe könnte sowohl durch ein im Vorfeld herzustellendes stärkeres Vertrauensverhältnis zwischen Teilnehmenden und Betreuungspersonen als auch durch eine noch stärkere Betonung darauf, dass nicht die Teilnehmenden selbst, sondern das Programm getestet wird, geschaffen werden. Auch verleitete die Betreuungssituation dazu, Unterstützung einzufordern: Wenn jemand als für die Unterstützung zuständig vorgestellt wird, wird diese Person ggf. gefragt, ohne selbst versucht zu haben, das Problem zu lösen. Des Weiteren wurde seitens der Betreuer und Betreuerinnen unterschiedlich schnell eingegriffen. Für eine Vergleichbarkeit wäre es hier sinnvoll gewesen, *den Betreuern und Betreuerinnen präzise Anweisungen unter Angabe einer festgelegten Zeitspanne vorzugeben, bevor sie eingreifen.*

(c) Fokussierte Interviews

In den Interviews wurde von den Probanden und Probandinnen nur selten konstruktive Kritik geübt. Erstens ist zu vermuten, dass aufgrund der sozialen Erwünschtheit vereinzelte Benutzbarkeits- und Akzeptanz-Probleme nicht angesprochen wurden. Der Interviewleitfaden wurde nach dem zweiten Evaluationszyklus dahingehend modifiziert, dass stärker Wert auf die Erfahrungsreflexion als auf die Softwarebewertung gelegt wurde. Zweitens sind die wenigen kritischen Anmerkungen möglicherweise darin begründet, dass sich die Probanden und Probandinnen nicht detailliert an die computerbasierte Diagnostik erinnern konnten. Um diese beiden vermuteten Ursachen für mangelnde Kritikformulierung zu reduzieren, wurden im dritten Evaluationszyklus während der Interviews Screenshots der Testumgebung als Anker bereitgehalten, um Erinnerungslücken zu schließen und einen Anknüpfungspunkt zu ermöglichen. Drittens ist möglich, dass es den Probanden und Probandinnen schwer fiel, sich sprachlich auszudrücken. Bei Detaillierungsfragen wurde deutlich, dass Probanden und Probandinnen die Frage nicht verstanden haben oder den Bezug zum Usability-Test nicht herstellen konnten. Eine mögliche Alternative zu fokussierten Interviews können Fokusgruppen darstellen, da sie sich unter Umständen unter „Ihresgleichen“ sicherer fühlen und sich ein Gespräch unter ihnen entwickelt. Die interviewende Person würde dadurch in den Hintergrund treten und es würden sich ggf. Hemmungen im Ausdruck reduzieren.

Retrospektiv könnte eine Verbindung zwischen den Methoden *Thinking Aloud* und der *Kommunikativen Validierung* eine passende Erhebungsform für die Zielgruppe und das Forschungsziel darstellen. Den Probanden und Probandinnen würden nachträglich Sequenzen ihres Handelns vorgespielt werden, um während dessen oder anschließend ihr Handeln zu begründen.

Schlussfolgernd sollten Interviewfragen an Personen, bei denen das Risiko der Akquieszenz und der sozialen Erwünschtheit besteht, offen formuliert werden an das Erfahrungswissen anknüpfen, ohne dabei zur expliziten Bewertung aufzufordern.

(d) Fragebögen zur Computererfahrung und -kompetenz

Der Einsatz der Fragebögen zur Computererfahrung und -kompetenz bedurfte keiner weiteren Explikationen oder Unterstützungen. Dies war vermutlich darin begründet, dass die Beantwortung der Fragen keine oder nur wenig Transferleistung erforderte. Insbesondere der Fragebogen zur Computererfahrung knüpfte an Erfahrungswissen an und setzte keine Erinnerung an einen konkreten Gegenstand voraus. Im Nachhinein betrachtet war allerdings problematisch, dass der Fragebogen zur Computerkompetenz erst nach dem Usability-Test eingesetzt wurde. Die Betreuer und Betreuerinnen wurden zwar instruiert, keine inhaltliche Unterstützung zu geben und es wurde davon ausgegangen, dass der Usability-Test keinen Lerneffekt hervorruft. Wie sich aber nach der Analyse der Videos herausstellte, unterstützen Betreuer und Betreuerinnen auch hinsichtlich der Funktion der Computermaus sowie der Hochstelltaste zur Großschreibung. Es kann somit nicht davon ausgegangen werden, dass das Vorwissen erfasst wurde. Schlussfolgernd sind *Befragungen zum Vorwissen vor der Interaktion durchzuführen. Sofern sie im Nachhinein durchgeführt werden sind mögliche Einfluss- bzw. Lernfaktoren zu kontrollieren.*

(e) User Experience Questionnaire (UEQ)

Nur zwei der sechs Skalen weisen eine mäßig zufriedenstellende Skalenkonsistenz auf. Die Ergebnisse zeigen, dass hauptsächlich die Mitte (4) und die beiden Extreme (1 und 7) angegeben wurden. Mögliche Gründe für die niedrigen Skalenkonsistenzen sind folgende:

- Die Verwendung von Gegensatzpaaren kann eine Überforderung darstellen, da zwei Wörter in einem Verhältnis zueinander interpretiert und auf das genutzte Instrument (otu.lea) transferiert werden müssen.
- Die siebenstufige Differenzierung der Antwortmöglichkeiten ist womöglich zu detailliert.
- Die Fragen zur Durchschaubarkeit können sowohl auf die Umgebung und Steuerung von otu.lea als auch auf die Itemschwierigkeit bezogen sein. Da die Personen sich zum Zeitpunkt der Testung auf unterschiedlichen Kompetenzniveaus befanden und unterschiedliche Niveaus bear-

beiteten, sind die Antworten beispielsweise zum Gegensatzpaar „einfach-kompliziert“ nicht aussagekräftig.

- Die Befragung wurde von unterschiedlichen Personen durchgeführt. Falls Probanden und Probandinnen Fragen zur Bedeutung der Gegensatzpaare gestellt haben, sind diese ggf. unterschiedlich durch die fragenden Personen operationalisiert worden.

Um diese möglichen Ursachen zu vermeiden und den Transfer der Attribute auf das Instrument zu erleichtern, könnte für jede Differenzierungsstufe ein auf das Instrument bezogenes Beispiel formuliert werden. Die Beispiele wären dann für alle Personen, welche die Befragung durchführen, einheitlich zu verwenden.

Möglicherweise könnte eine leichte Modifikation Abhilfe schaffen: Um das Gefühl einer Notengebung zu vermeiden könnten die Zahlen weggelassen werden. Um Suggestion zu vermeiden könnten die Probanden und Probandinnen selbst das zutreffende Attribut ankreuzen. Obwohl die Version „Simple Language“ verwendet wurde, scheint der UEQ wenig geeignet für den Einsatz bei dieser Zielgruppe. Demnach sind *Befragungen, die eine Differenzierung ohne qualifizierende Beispiele erfordern, entweder zu vermeiden oder es sind qualifizierende Beispiele bzw. Operationalisierungen der Abstufungen vorzunehmen.*

Zusammenfassend scheinen interaktionszentrierte bzw. benutzer-/benutzerinnenzentrierte Messansätze, die entweder eine hohe Transferleistung (a), eine Abstraktion (c) oder einen zeitlich verzögerten Bezug (d) erfordern, für die Zielgruppe nur bedingt geeignet. Die Befragungen zur Computererfahrung und -kompetenz (e) liefern hilfreiche Erkenntnisse und können in weiteren Software-Entwicklungen (sowohl für Diagnose- als auch Lernprogramme) berücksichtigt werden. Der interaktionszentrierte Messansatz bzw. „reines“ Usability-Testing (b) am Computer ist sehr aufschlussreich und lässt umfangreiche Schlussfolgerungen bezüglich zielgruppenspezifischer Usability-Anforderungen zu.

9.3 Stichprobenauswahl

Zur Stichprobenwahl ist anzumerken, dass durch die ausschließliche Zugänglichkeit zur Zielgruppe über Alphabetisierungskurse ein positiver Selektionseffekt stattgefunden haben kann. Alle befragten Personen nahmen zum Zeitpunkt der Studien bereits an Alphabetisierungskursen teil, d. h. sie haben bereits Lerneinheiten durchlaufen. Auch handelt es sich höchstwahrscheinlich um Personen, die lernmotiviert sind, da die meisten von ihnen freiwillig die Kurse besuchen. Zu vermuten ist daher, dass an der Studie tendenziell die kompetenzstärkeren und stärker motivierten Personen im Vergleich zur Gesamtheit der funktionalen Analphabeten und Analphabetinnen in Deutschland teilgenommen haben.

Der Zugang zu der Zielgruppe in der Größenordnung war allerdings nur über Alphabetisierungskurse möglich, da funktionale Analphabeten und Analphabetinnen, die nicht an Alphabetisierungskursen teilnehmen, schwer auffindbar sind. Die Ergebnisse sind daher auch vor dem Hintergrund der Stichprobenauswahl zu bewerten. Allerdings ist ebenfalls anzunehmen, dass vor allem diejenigen, die sich entschieden haben, ihre Literalität zu verbessern und Alphabetisierungskurse besuchen, primär diejenigen sind, die otu.lea nutzen. Zudem haben zwei Personen an der Studie teilgenommen, die vorher nie mit dem Computer gearbeitet haben, so dass zumindest hinsichtlich der Computererfahrung und -kompetenz davon ausgegangen werden kann, die Grundgesamtheit zu repräsentieren.

10 Forschungsdesiderata und Ausblick

Die Forschungsbemühungen im Bereich der Alphabetisierung im Allgemeinen und computerbasierten Diagnostik im Grundbildungsbereich im Besonderen sind erst seit wenigen Jahren intensiviert worden. Für zukünftige Forschungsbemühungen ergeben sich aus den hier dargestellten Ergebnissen und Überlegungen a) Forschungsdesiderata für den Literalitätsdiskurs und die computerbasierte Diagnostik im Grundbildungsbereich im Allgemeinen, b) Forschungsdesiderata auf der Grundlage der Forschungsergebnisse sowie c) Entwicklungspotenziale für das Online-Diagnoseinstrument *otu.lea*. Diese werden im Folgenden erläutert um mit einem Ausblick und einer kurzen Zusammenfassung zu schließen.

a) Forschungsdesiderata für den Literalitätsdiskurs und die computerbasierte Diagnostik im Grundbildungsbereich

Unter Berücksichtigung des Diskurses zum Literalitätsbegriff ist zu überlegen, in wie weit die ICT-Literacy mit in das zu prüfende Konstrukt der Literalität aufgenommen werden sollte. Genuin bietet für die Überprüfung der ICT-Literacy ein computerbasiertes Diagnoseinstrument ideale Voraussetzungen. Hierfür könnten auch Entwicklungen für die Messung einer technologiespezifischen komplexen Problemlösefähigkeit berücksichtigt werden.

Auf der Ebene der Methodendiskussion ist weiterhin zu explorieren, welche Methoden unter Berücksichtigung der zielgruppenspezifischen Charakteristika (Testangst, Schamgefühl) einsetzbar sind, um Nervosität sowie das Risiko der Zustimmungstendenz und der sozialen Erwünschtheit zu reduzieren. Dafür wäre es sinnvoll, Erfahrungen mit der Zielgruppe sowie Ergebnisse von Evaluationsstudien inklusive einer Methodenreflexion/-diskussion der Fachwelt zu kommunizieren und als Diskussionsgegenstand zur Verfügung zu stellen – wie es auch grundsätzlich im DBR vorgesehen ist. Erst unter der Berücksichtigung der zielgruppenspezifischen Herausforderungen und somit der Gestaltung sowie des sinnvollen Einsatzes der Erhebungsmethoden sind aussagekräftige Ergebnisse zu erwarten.

Hinsichtlich der computerbasierten Diagnostik ist – neben weiteren Forschungsbemühungen zur Usability – die Äquivalenzproblematik intensiver zu thematisieren und zu untersuchen. Einerseits könnte eine Äquivalenzprüfung vorgenommen werden, indem Ergebnisse mit dem papierbasierten und dem computerbasierten Instrument vorgenommen werden (vgl. Ergebnisdiskussion zur Computererfahrung in Kap. 8.3.6). Andererseits ist anzunehmen, dass ein Unterschied zwischen dem Tippen mit der Tastatur und dem selbstständigem Schreiben mit einem Stift (der eigenständigen Konstruktion von Buchstaben)

besteht. Es ist bisher aber nicht erforscht, in wie weit das Tippen mit der Tastatur den Konstruktionsprozess von Wörtern unterstützt und daher ggf. die Äquivalenz zwischen der computerbasierten und der papierbasierten Diagnostik beeinträchtigt. Möglich wäre allerdings auch, dass das zwischenzeitliche Tippen auf einer Tastatur den Lernprozess unterstützt.

Vor dem Hintergrund dieser Überlegungen zur computerbasierten Diagnostik schließt sich die Frage an, ob Usability-Richtlinien speziell für das computerbasierte Testen einzuführen sind. Richtlinien für eine *Test -Usability* (vgl. Koppel, 2011) würden den Einsatzkontext sowie die Testmodalität stärker berücksichtigen, indem insbesondere die Vermeidung des Einflusses der Computererfahrung und -kompetenz auf das Testergebnis priorisiert werden würde.

Zudem wurde in der Alphabetisierungsforschung bisher wenig beachtet, welchen Einfluss die individuellen Motive auf den Lernprozess der funktionalen Analphabeten und Analphabetinnen haben. Es ist bekannt, dass funktionale Analphabeten und Analphabetinnen das Internet und Smartphones nutzen. Das Ziel, sich sicherer und wissender im Internet zu bewegen ist für Internet- und Smartphone-nutzer/-nutzerinnen oftmals ein Anreiz, ihre literalen Kompetenzen weiterzuentwickeln. Das Internet kann durch seine unendlichen Informationsquellen zudem ein Literalitätsförderndes Instrument darstellen. Beispielweise bietet es die Möglichkeit, interessenbezogene Texte auszuwählen (Liedtexte, Rezepte etc.) und sich diese ohne Zeitdruck zu erschließen. So kann ggf. das Erschließen und Verstehen bestimmter Informationen ebenfalls ein Anreiz sein, die eigenen Literalitätsbemühungen zu steigern. Es wäre sinnvoll, weitere Untersuchungen zu Motiven für das Lesen- und Schreibenlernen anzustellen, um mehr Personen zu erreichen und die Literalisierung motivierend zu gestalten.

b) Forschungsdesiderata auf der Grundlage der Forschungsergebnisse

Im Folgenden werden Forschungsdesiderata formuliert, die sich einerseits auf eine gebrauchstaugliche Gestaltung einer Online-Diagnostik für den Grundbildungsbereich beziehen und andererseits mit dem Online-Diagnose-Instrument *otu.lea* aufgegriffen werden können: *otu.lea* bietet die Möglichkeit, Untersuchungen mit größeren Stichproben durchzuführen, aus denen weitere Überprüfung und Konkretisierung der Heuristiken und schließlich die Generierung von allgemeingültigen Ergebnissen abgeleitet werden können. In Anbetracht der Forschungsfragen und Ergebnisse des hier vorgestellten Forschungsvorhabens ist in weiteren Untersuchungen die Fokussierung folgender Themenbereiche sinnvoll:

- Für eine Generalisierbarkeit wären die Heuristiken mit einer größeren Stichprobe zu überprüfen. Als darauf folgenden Schritt wäre es möglich, die Heuristiken zu konkretisieren und wiederum als Basis für die

Entwicklung von Guidelines (detaillierten Gestaltungsrichtlinien) heranzuziehen.

- Es wäre zu überprüfen, welche der WCAG-Richtlinien auch für funktionale Analphabeten und Analphabetinnen oder gerade nicht für funktionale Analphabeten und Analphabetinnen relevant sind.
- Darüber hinaus wären weitere Untersuchungen zum *Cognitive Load* mit – in diesem Kontext bisher wenig untersuchten – Personen im Grundbildungsbereich durchzuführen und in die Überlegungen zur gebrauchstauglichen Gestaltung von Interfaces einzubeziehen. Dabei wären auch die Einbindung von Multimedia explizit zu berücksichtigen; es wäre zu untersuchen, welche Faktoren für den *Cognitive Load* und insbesondere des *Extraneous Load* reduzierend sind. Zudem ist beispielsweise fraglich, ob ein Grenzbereich zwischen einer *Overcomplication* und *Oversimplification* identifiziert werden kann, der ein ideales Anspruchsniveau für Personen im Grundbildungsbereich darstellt. Hier wäre auch der Einsatz adaptiver Tests sinnvoll.

Dass die Entwicklung von computerbasierten (Förder-)Diagnoseinstrumenten in der Grundbildung Erwachsener erst am Anfang steht ist vorwiegend in einer, bis vor wenigen Jahren noch, geringen Aufmerksamkeit für das Thema Alphabetisierung begründet. Zudem steht hinter der Alphabetisierung keine ressourcenstarke Lobby, durch die Investitionen getätigt werden. Die Beschäftigten in diesem Bereich sind meistens in prekären Arbeitsverhältnissen auf Honorarbasis; finanzielle Unterstützung gibt es primär vom Bund und den Ländern. Unter diesen Voraussetzungen ist fraglich, in wie weit Ressourcen akquiriert werden können, um otu.lea weiterzuentwickeln und eine Dissemination in der Alphabetisierungspraxis zu realisieren.

c) Entwicklungspotenziale für das Online-Diagnoseinstrument otu.lea

Auch wenn das Online-Diagnose-Instrument für die Alphabetisierung in dem gegenwärtigen Entwicklungsstand ein gewinnbringendes, sinnvoll einsatzfähiges und qualitativ hochwertiges Instrument darstellt, bestehen Weiterentwicklungspotenziale. Auf der Basis der Softwarearchitektur sind weitere Einsatzfelder und Weiterentwicklungen denkbar und umsetzbar¹²⁸:

- otu.lea kann im beruflichen Übergangssystem, in der Berufsvorbereitung, aber auch in Ausbildungskontexten genutzt werden, um auf höheren Kompetenzstufen den Kompetenzentwicklungsbedarf von (Berufs-)Schüler und -Schülerinnen sowie Auszubildenden festzustellen. Insbesondere im Berufsbildungskontext sind wenig zeitliche Ressourcen vorhanden, um speziell die literalen und mathematischen Grundkompeten-

¹²⁸ Die Weiterentwicklungspotenziale wurden bereits in Wolf und Koppel (2014) publiziert.

zen zu fördern. Mit der ausführlichen, computerbasierten Diagnostik und der Berücksichtigung der Kompetenzentwicklung wäre der Lernbedarf schnell identifizierbar.

- Durch die Integration der lea.-Lernmaterialien können neben der Diagnostik auch Übungsmaterialien bereitgestellt werden, die direkt an die Kompetenzfeststellung anknüpfen. Somit bestünde nach der Diagnostik die Möglichkeit, die zum Kompetenzstand passenden Übungsaufgaben vorzuschlagen. Diese Verzahnung würde die Diagnose- und Fördermaterialien zu einem effektiven und effizienten Förderdiagnoseinstrument vereinen.
- Um der Entwicklung der zunehmenden Nutzung von Smartphones sowie Tablets standzuhalten und zudem eine zeit- und ortsunabhängige Nutzung von otu.lea zu ermöglichen, wäre otu.lea für mobile Endgeräte aufzubereiten. Dies würde eine stärkere Privatsphäre ermöglichen, da kein öffentlicher oder familiär geteilter Desktoprechner genutzt werden müsste.
- Im Kontext der beruflichen Bildung ist die Entwicklung berufsspezifischer Aufgaben zur Förderung und Diagnostik der Literalität bzw. Fachsprachlichkeit in einem Fachkontext wie z. B. einem Ausbildungsberuf möglich und sinnvoll. Dies würde auch den Bedarfen entsprechen, effizient und kurzzeitig fehlende literale Grundbildung nachzuholen sowie authentische Inhalte berufsrelevanter Kontexte in Lern- und Diagnosematerialien in der beruflichen Bildung einzubinden.

Weitere Entwicklungspotenziale gehen mit der Technikentwicklung einher. So bestehen bereits Möglichkeiten, auch offene Itemformate automatisiert auszuwerten (vgl. z. B. Burstein u. a., 1998; Streeter u. a., 2011; Sukkariéh & Stoyanchev, 2009). Dies hat in der Alphabetisierung eine besondere Relevanz, da in der Literalitätsförderung dem freien Schreiben eine besondere Bedeutung beigemessen wird. Bisher ist die automatisierte Auswertung offener Antwortformate jedoch sehr zeit- und kostenintensiv, daher bedarf es auch hier weiterer intensiver Forschungsbemühungen.

Ausblickend ist für die nahe Zukunft nicht zu erwarten, dass die Anzahl funktionaler Analphabeten und Analphabetinnen merklich sinken wird. Dafür sprechen einerseits die Ergebnisse der Studie Alpha-Panel, die auf einen nur begrenzten Lernerfolg in Alphabetisierungskursen hinweist (Rosenblatt & Lehmann, 2013b). Andererseits lassen die PISA-Ergebnisse erwarten, dass trotz einer Verbesserung der Lesekompetenz seit der ersten PISA-Studie 2000 jährlich immer noch ca. 18% eines Altersjahrganges mit Lese-, Schreib- und Rechenschwächen zu den bereits bestehenden funktionalen Analphabeten und Analphabetinnen hinzukommen. Durch die Weiterentwicklung und Verbreitung digitaler

Spracherkennung, der Möglichkeit, Bilder in Kurznachrichten einzubinden sowie der automatisierten Bild- und Worterkennung ist sogar zu vermuten, dass die Anzahl steigen wird. Denn die Notwendigkeit, eigenständig Texte schreiben zu können wird mit zunehmender Nutzung solcher Technologien sinken.

Zusammenfassend bedarf es weiterhin umfangreicher Alphabetisierungsbemühungen, der Beendigung prekärer Beschäftigungsverhältnisse, der Evaluationstätigkeiten für die Kursarbeit in der Alphabetisierung (beispielsweise durch Lernstandsmessungen mit Hilfe von otu.lea) sowie der Weiterentwicklung von computerbasierter Diagnostik in der Grundbindung, um eine Qualitätssteigerung in der Kursarbeit und somit auch eine Reduzierung des funktionalen Analphabetismus zu erreichen. Mit der Entwicklung und Evaluation von otu.lea wurde ein Beitrag zur Alphabetisierung geleistet: Durch die systematische Überprüfung und Weiterentwicklung konnte sowohl ein theoretischer als auch ein praktischer Output generiert werden. Die Online-Diagnostik bietet die Möglichkeit einer anonymen Nutzung sowie den Einsatz in Alphabetisierungskursen und begegnet damit Herausforderungen in der Alphabetisierungsarbeit. Bis zum Herbst des Jahres 2014 wurde otu.lea fortwährend weiterentwickelt und evaluiert. Inzwischen ist eine vollwertige Version unter www.otulea.uni-bremen.de abrufbar.

Literatur

- Ainley, M. (2006). Connecting with Learning: Motivation, Affect and Cognition in Interest Processes. *Educational Psychology Review*, 18(4), 391–405.
- Akker, J. V. den, Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational Design Research*. Routledge.
- Ally, M. (2004). Foundations of educational theory for online learning. In T. Anderson & F. Elloumi (Hrsg.), *Theory and Practice of Online Learning* (S. 3–31). Athabasca: Athabasca University.
- Anderson, J. R. (2007). *Kognitive Psychologie*. Heidelberg: Spektrum, Akad. Verl.
- Anderson, J. R. (2013). *Kognitive Psychologie* (7. Aufl.). Berlin, Heidelberg: Springer.
- Anderson, T., & Shattuck, J. (2012). Design-Based Research: A Decade of Progress in Education Research? *Educational Researcher*, (41), 16–25.
- ANLCI - French National Agency for the Fight against Illiteracy (Hrsg.). (2009). Basic Skills Competence Evaluation Policies in Several European Countries.
- Anslinger, E., & Quante-Brandt, E. (2010). Grundbildung am Übergang Schule-Beruf und die Bedeutung der individuellen Kompetenzentwicklung. *bwp@Berufs- und Wirtschaftspädagogik - online*, (18), 1–19.
- ARD/ZDF-Medienkommission (Hrsg.). (2014). ARD/ZDF-Online-studie 2014.
- Baddeley, A. D. (1976). *The psychology of memory*. New York: Basic Books.

- Ball, S. (2009). Accessibility in e-assessment. *Assessment & Evaluation in Higher Education*, 34(3), 293–303.
- Barab, S. A., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of The Learning Sciences*, 13(1), 1–14.
- Belusa, A., & Eberwein, H. (1997). Förderdiagnostik – Eine andere Sichtweise diagnostischen Handelns. In *Handbuch Integrationspädagogik. Kinder mit und ohne Behinderung lernen gemeinsam* (S. 260–268). Weinheim und Basel: Beltz.
- Bennett, R. E. (1993). On the Meaning of Constructed Response. In R. E. Bennett & W. C. Ward (Hrsg.), *Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment* (S. 1–28). Hillsdale, New Jersey: Lawrence Erlbaum.
- Bennett, R. E. (1998). Reinventing Assessment: Speculations on the Future of Large-Scale Educational Testing. ETS Policy Information Center Report.
- Bennett, R. E. (2001). How the Internet Will Help Large-Scale Assessment Reinvent Itself. *Education Policy Analysis Archives*, 9(5).
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. *The Journal of Technology, Learning and Assessment*, 6(9).
- Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring Problem Solving with Technology: A Demonstration Study for NAEP. *The Journal of Technology, Learning and Assessment*, 8(8).

- Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society*, 9(3), 321–327.
- Bergius, R. (2004). Aufmerksamkeit. In H. O. Häcker & K.-H. Stapf (Hrsg.), *Psychologisches Wörterbuch* (14. Aufl., S. 84–85). Bern: Hans Huber.
- Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is Usability? Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals, 651–655.
- BMBF. (2014). Nationale Strategie für Alphabetisierung und Grundbildung Erwachsener - Bildung - BMBF. Abgerufen 25. September 2014, von <http://www.bmbf.de/de/426.php>
- Bogner, A., Littig, B., & Menz, W. (2005). *Das Experteninterview: Theorie, Methode, Anwendung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bond, T. G., & Fox, C. M. (2001). Applying the Rasch Model: Fundamental Measurement in the Human Sciences (1. Aufl.). Psychology Press.
- Bonna, F., & Nienkämper, B. (2011). Diagnostik nicht ohne Kursleiter/innen - Begründungen, Konzeption und erste Ergebnisse einer Befragung an Volkshochschulen. In B. Egloff & A. Grotlüschen (Hrsg.), *Alphabetisierung und Grundbildung* (S. 35–52). Münster: Waxmann.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler* (4., überarb. Aufl.). Springer Berlin Heidelberg.
- Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: what are they and what are their benefits? *Assessment & Evaluation in Higher Education*, 34(3), 305–319.
- Braude, E. J., & Bernstein, M. E. (2011). *Software engineering: modern approaches*. Hoboken, NJ: Wiley.

- Breitenbach, E. (2007). *Förderdiagnostik*. Würzburg: Edition Bentheim.
- Brennan, R. L., & Prediger, D. J. (1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*, 41(3), 687–699.
- Briggs, L. J., & Gagné, R. M. (1974). *Principles of instructional design*. New York: Holt, Rinehart and Winston.
- Brigitte, o. A. (2004). Mein erstes Beratungsgespräch bei der Volkshochschule. *Alfa-Forum*, 17(56), 19–22.
- Brooks, G., Davies, R., Ducke, L., Hutchison, D., Kendall, S., & Wilkin, A. (2001). Progress in Adult Literacy: Do Learners learn? (Basic Skills Agency, London & National Foundation for Educational Research, Hrsg.).
- Brooks, G., Heath, K., & Pollard, A. (2005). Assessing adult literacy and numeracy: a review of assessment instruments. London: NRDC.
- Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *The Journal of the Learning Sciences*, 2(2), 141–178.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18(1), 32–42.
- Brügelmann, H. (2000). *Kinder auf dem Weg zur Schrift: eine Fibel für Lehrer und Laien*. Bottighofen: Libelle-Verl.
- Bruner, J. (1994). *Four ways to make meaning*. Gehalten auf der AERA Conference, New Orleans.
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of Cognitive Load in Multimedia Learning with Dual-Task Methodology: Auditory Load and Modality Effects. *Instrucional Science*, (32), 115–132.

-
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring Cognitive Load. In J. L. Plass, R. Moreno, & R. Brünken (Hrsg.), *Cognitive Load Theory* (S. 181–202). Cambridge University Press.
- Brünken, R., Steinbacher, S., Schnotz, W., & Leutner, D. (2001). Mentale Modelle und Effekte der Präsentations- und Abrufkodalität beim Lernen mit Multimedia. *Zeitschrift für Pädagogische Psychologie*, 15(1), 16–27.
- Büning, H. (1991). *Robuste und adaptive Tests*. Berlin u.a.: de Gruyter.
- Burkhardt, H., & Schoenfeld, A. H. (2003). Improving Educational Research: Toward a More Useful, More Influential, and Better-Funded Enterprise. *Educational Researcher*, 32(9), 3–14.
- Burstein, J., Kukich, K., Wolff, S., Lut, C., Chodorow, M., Braden-harder, L., & Harris, M. D. (1998). Automated Scoring Using A Hybrid Feature Identification Technique. In *In the Proceedings of the Annual Meeting of the Association of Computational Linguistics* (S. 206–210).
- Caldwell, B., Cooper, M., Reid, L. G., & Vanderheiden, G. (2008). Web Content Accessibility Guidelines (WCAG) 2.0. W3C.
- Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*, 8(4), 293–332.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, 32(1), 9–13.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Hrsg.), *New Directions in Educational Technology*. Berlin: Springer.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*, 13(1), 15–42.

- Colombo-Mendoza, L. O., Alor-Hernandez, G., & Rodriguez-Gonzalez, A. (2012). A novel approach for generating multi-device Rich Internet Applications. In *2012 22nd International Conference on Electrical Communications and Computers (CONIELECOMP)* (S. 361–367).
- Cook, A. E., Zheng, R. Z., & Blaz, J. W. (2009). Measurement of Cognitive Load During Multimedia Learning Activities. In *Cognitive Effects of Multimedia Learning* (S. 34–50). London: Idea Group Publishing.
- Crawford, C. (2004). *Chris Crawford on Interactive Storytelling* (Auflage: Revised.). Berkeley, CA: New Riders Publ.
- Creswell, J. W., & Clark, V. L. P. (2010). *Designing and Conducting Mixed Methods Research* (Auflage: 2nd edition.). Los Angeles: Sage Publications Ltd.
- Crisp, G. (2007). *The e-assessment handbook*. [London]; New York: Continuum.
- Csapó, B., Ainley, J., Bennett, R. E., Latour, T., & Law, N. (2012). Technological Issues for Computer-Based Assessment. In P. Griffin, B. McGaw, & E. Care (Hrsg.), *Assessment and Teaching of 21st Century Skills* (S. 143–230). Springer Netherlands.
- Cummins, J., Ardeshiri, M., & Cohen, S. (2008). Computer-Supported Scaffolding of Literacy Development. *Pedagogies: An International Journal*, 3(1), 4–21.
- Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Manage. Sci.*, 32(5), 554–571.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5–6), 523–545.
- Deci, E. L., & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Psychologie*, 39, 223–238.

- Dennis, A. R., & Valacich, J. S. (1999). Rethinking Media Richness: Towards a Theory of Media Synchronicity. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences-Volume 1 - Volume 1* (S. 1017–). Washington, DC, USA: IEEE Computer Society.
- Department for Business, Innovation and Skills. (2011). 2011 - Skills for Life Survey: Headline findings. Abgerufen von <http://www.nationalnumeracy.org.uk/resources/12/index.html>
- Die Senatorin für Bildung Wissenschaft und Gesundheit (Hrsg.). (2011). Bericht zur vorschulischen Sprachstandsfeststellung 2011. Abgerufen von <http://www.bildung.bremen.de/sixcms/media.php/13/TOP+5.pdf>
- DIN, (Deutsches Institut für Normung) (Hrsg.). (2006). DIN EN ISO 9241-110: Europäisches Komitee für Normung, Ergonomie der Mensch-System-Interaktion. Teil 110: Grundsätze der Dialoggestaltung.
- DIN Deutsches Institut für Normung e. V. (Hrsg.). (2008). Grundsätze der Dialoggestaltung: (ISO 9241-110:2006) ; Deutsche Fassung EN ISO 9241-110:2006. Berlin: Beuth.
- DIN EN ISO. (2009). DIN EN ISO 9241-20:2009-08 - Ergonomie der Mensch-System-Interaktion - Teil 20: Leitlinien für die Zugänglichkeit der Geräte und Dienste in der Informations- und Kommunikationstechnologie.
- DIN EN ISO. (2011). Grundsätzliche Empfehlungen für Produkt- und Prozessgestaltung (1. Aufl.). Berlin: Beuth.
- Dluzak, C., Heinemann, A., & Grotluschen, A. (2009). Vorschlag für neue „Alpha-Levels“. Mehr untere Sprossen für die Leiter. *DIE Zeitschrift für Erwachsenenbildung*, 16(1), 34–36.
- Döbert, M., Hubertus, P., & Nickel, S. (2000). *Ihr Kreuz ist die Schrift: Analphabetismus und Alphabetisierung in Deutschland* (1. Aufl.). Bundesverband Alphabetisierung.

- Dupuis, G., & Kerkhoff, W. (1992). Enzyklopädie der Sonderpädagogik, der Heilpädagogik und ihrer Nachbargebiete. Berlin: Wissenschaftsverlag Spiess.
- Edelson, D. C. (2002). Design Research: What We Learn When We Engage in Design. *The Journal of the Learning Sciences*, 11(1), 105–121.
- Eggert, D. (1997). Von den Stärken ausgehen ...: individuelle Entwicklungspläne in der Lernförderungsdiagnostik; ein Plädoyer für andere Denkgewohnheiten und eine veränderte Praxis; [mit Formular-Kopiervorlagen]. Dortmund: Borgmann.
- Egloff, B. (1997). Biographische Muster „funktionaler Alphabeten“. Eine biographieanalytische Studie zu Entstehungsbedingungen und Bewältigungsstrategien von „funktionalem Alphabetismus“, (Deutsches Institut für Erwachsenenbildung).
- Engel, N. (2008). Förderdiagnostik in der Alphabetisierung: Eine empirische Untersuchung zur Schreibprozessdiagnose in Alphabetisierungskursen Niedersachsens. Stuttgart: Ibidem.
- Fischer, F., Bouillion, L., & Mandl, H. (2003). Bridging theory and practice in learning environment research - Scientific principles in pasteur's quadrant. *International Journal of Educational Policy Research, and Practice*, 4(1), 147–170.
- Fisk, A. D., Rogers, W. A., Charness, N., Sharit, J., & Czaja, S. J. (2009). *Designing for Older Adults: Principles and Creative Human Factors Approaches* (0002 Aufl.). Crc Pr Inc.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613–619.
- Fowler, A. M. L. (2008). Providing effective feedback on a wholephrase input in computer-assisted language learning. Computing Department, University of Kent, UK.

- Friebertshäuser, B., & Langer, A. (2010). Interviewformen und Interviewpraxis. In B. Friebertshäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 437–455). Weinheim [u.a.]: Juventa.
- Friedrich, H. F., & Mandl, H. (1997). Analyse und Förderung selbstgesteuerten Lernens. In *Enzyklopädie der Psychologie* (S. 237–293). Göttingen: Hogrefe.
- Frieling, G., & Rustemeyer, A. (2011). Die Volkshochschulen, der Forschungstransfer und das Recht auf Alphabetisierung. In B. Egloff & A. Grotlüschen (Hrsg.), *Forschen im Feld der Alphabetisierung und Grundbildung: Ein Werkstattbuch*. Münster; New York, NY; München; Berlin: Waxmann.
- Frith, U. (1985). Beneath the Surface of Developmental Dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Hrsg.), *Surface Dyslexia, Neuropsychological and Cognitive Studies of Phonological Reading* (S. 301–330). London: Erlbaum.
- Füssenich, I. (2004). Diagnostik - nicht schon wieder, oder? *Alfa-Forum*, 17(56), 10–11.
- Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the Consequential Validity of New Modes of Assessment: The Influence of Assessment on Learning, Including Pre-, Post-, and True Assessment Effects. In M. Segers, F. Dochy, & E. Cascallar (Hrsg.), *Optimising New Modes of Assessment: In Search of Qualities and Standards* (S. 37–54). Springer Netherlands.
- Glaser, B. G., Strauss, A. L., & Paul, A. T. (2010). *Grounded Theory Strategien qualitativer Forschung*. Bern: Huber.
- Glöckner, C. (2013). Das Bildungswesen in China. In C. Adick (Hrsg.), *Bildungsentwicklungen und Schulsysteme in Afrika, Asien, Lateinamerika und der Karibik* (S. 191–212). Münster, Westf.: Waxmann.

- Gould, J. D. (1985). Design for Usability: Key Principles and What Designers Think. *Commun. ACM*, 28, 300–311.
- Graf, R., Zimbardo, P. G., & Gerrig, R. J. (2007). *Psychologie*. München u. a.: Pearson-Studium.
- Gravenmeijer, K., & Cobb, P. (2006). Design Research from a technology perspective. In J. V. den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Hrsg.), *Educational Design Research* (S. 45–85). New York: Routledge.
- Grotlüschen, A. (Hrsg.). (2010). *lea-Literalitätsentwicklung von Arbeitskräften - Diagnose*. Münster: Waxmann.
- Grotlüschen, A. (2011). Zur Auflösung von Mythen. In *Literalitätsentwicklung von Arbeitskräften* (S. 12–39). Münster: Waxmann.
- Grotlüschen, A. (2012). Literalität und Erwerbstätigkeit. In A. Grotlüschen & W. Riekmann (Hrsg.), *Funktionaler Analphabetismus in Deutschland: Ergebnisse der ersten leo. - Level-One-Studie* (S. 135–165). Münster: Waxmann.
- Grotlüschen, A., & Heinemann, A. (2011). Ergebnisse der Schwierigkeitsbestimmung förderdiagnostischer Aufgaben. In A. Grotlüschen, R. Kretschmann, E. Quante-Brandt, & K. D. Wolf (Hrsg.), *Literalitätsentwicklung von Arbeitskräften* (S. 108–121). Münster: Waxmann.
- Grotlüschen, A., & Riekmann, W. (2011). leo.-Level-One-Studie. Presseheft. Abgerufen von <http://blogs.epb.uni-hamburg.de/leo/>
- Grotlüschen, A., & Riekmann, W. (2012). *Funktionaler Analphabetismus in Deutschland. Ergebnisse der ersten leo. - Level-One Studie*. Waxmann Verlag.
- Grotlüschen, A., Riekmann, W., & Buddeberg, K. (2012). Hauptergebnisse der leo.-level One Studie. In *Funktionaler Analphabetismus in Deutschland. Ergebnisse der ersten leo. - Level-One Studie* (S. 13–53). Münster: Waxmann.

- Gulikers, J., Bastiaens, T., Kirschner, P. A., & Kester, L. (2008). Authenticity is in the Eye of the Beholder: Student and Teacher Perceptions of Assessment Authenticity. *Journal of Vocational Education & Training*, 60(4), 401–412.
- Hackos, J. T. (1998). *User and task analysis for interface design*. New York u. a.: Wiley.
- Hadwin, A. F., Winne, P. H., & Nesbit, J. C. (2005). Roles for software technologies in advancing research and theory in educational psychology. *The British Journal of Educational Psychology*, 75(Pt 1), 1–24.
- Haladyna, T. M. (2004). Developing and validating multiple-choice test items. Routledge.
- Hare, A. P. (1960). Interview Responses: Personality or Conformity? *Public Opinion Quarterly*, 24, 679–685.
- Hasebrook, J. (2006). Aptitude-Treatment-Interaktion. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (Bd. 3, S. 20–26). Weinheim: Beltz.
- Hassenzahl, M., Burmester, M., & Koller, F. (2003, Januar 1). AttrakDiff: ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität [Konferenz- oder Workshop-Beitrag]. Abgerufen 15. März 2013, von <http://tubiblio.ulb.tu-darmstadt.de/21687/>
- Heinemann, A. (2011). Alpha-Levels, lea.-Universum, Item-Entwicklung und Feldzugang. In A. Grotluschen, R. Kretschmann, E. Quante-Brandt, & K. D. Wolf (Hrsg.), *Literalitätsentwicklung von Arbeitskräften* (S. 86–102). Münster: Waxmann.
- Helfferrich, C. (2004). *Die qualität qualitativer Daten* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Herczeg, M. (2009). Software-Ergonomie Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme. München: Oldenbourg.

- Herrington, J., McKenney, S., Reeves, T., & Oliver, R. (2007). Design-based research and doctoral students: Guidelines for preparing a dissertation proposal. In C. Montgomerie & J. Seale (Hrsg.), (Bd. 2007, S. 4089–4097). *Gehalten auf der Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Chesapeake.
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Faculty of Education - Papers (Archive)*.
- Hertzum, M. (2010). Images of Usability. *International Journal of Human-Computer Interaction*, 26(6), 567–600.
- Hertzum, M., & Jacobsen, N. E. (2003). The evaluator effect: a chilling fact about usability evaluation methods. *Int. Journal of Human-Computer Interaction*.
- Higher Education Funding Council. (1999). Guidance on submissions research assessment exercise, Paragraph 1.12. (Higher Education & Funding Council for England and Wales, Hrsg.).
- Hornecker, E. (2004). *Tangible User Interfaces als kooperationsunterstützendes Medium*.
- Huntemann, H., & Reichart, E. (2011). *Volkshochschul-Statistik 2010.: 49. Folge, Arbeitsjahr 2010. Stand Informationen: Oktober 2011*. Deutsches Institut für Erwachsenenbildung - Leibniz-Zentrum für Lebenslanges Lernen, Bonn. Abgerufen von <http://www.die-bonn.de/doks/2011-volkshochschule-statistik-01.pdf>
- Huntemann, H., & Reichart, E. (2013). *Volkshochschul-Statistik 51. Folge, Arbeitsjahr 2012*. Deutsches Institut für Erwachsenenbildung - Leibniz-Zentrum für Lebenslanges Lernen, Bonn. Abgerufen von <http://www.die-bonn.de/doks/2011-volkshochschule-statistik-01.pdf>
- IBM. (2007). *Virtual Worlds, Real Leaders: Online games put the future of business leadership on display*.

- Ingenkamp, K., & Lissmann, U. (2008). *Lehrbuch der pädagogischen Diagnostik*. Weinheim; Basel: Beltz.
- International Organization for Standardization, & International Electrotechnical Commission (Hrsg.). (2012). Information technology — W3C Web Content Accessibility Guidelines (WCAG) 2.0. ISO.
- Jacobs, B. (2008). Was wissen wir über die Lernwirksamkeit von Aufgabenstellung und Feedback. In Aufgaben als Katalysatoren von Lernprozessen: eine zentrale Komponente organisierten Lehrens und Lernens aus der Sicht von Lernforschung, allgemeiner Didaktik und Fachdidaktik (S. 99–114). Münster: Waxmann.
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, 1(2), 112–133.
- Jordan, S. (2009). Assessment for learning: pushing the boundaries of computer-based assessment. *Practitioner Research in Higher Education*, 3(1), 11–19.
- Jurecka, A., & Hartig, J. (2007). Computer- und netzwerkbasierendes Assessment. *Möglichkeiten und Voraussetzungen technologisierter Kompetenzdiagnostik*, (Bundesministerium für Bildung und Forschung), 37–48.
- Kantowitz, B. H. (1987). Mental Workload. In P. A. Hancock (Hrsg.), *Human Factors Psychology* (1 edition., S. 81–121). Amsterdam ; New York : New York, N.Y., U.S.A: North Holland.
- Ketelhut, D. J., & Nelson, B. C. (2010). Designing for real-world scientific inquiry in virtual environments. *Educational Research*, 52(2), 151–167.
- Ketelhut, D. J., Nelson, B., Clarke, J., & Dede, C. (2010). A Multi-user virtual environment for building higher order inquiry skills in science. *British Journal of Educational Technology*, 41(1), 56–68.

- Kim, M. C., & Hannafin, M. J. (2011). Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice. *Computers & Education*, 56(2), 403–417.
- Kirschner, F., Kester, L., & Corbalan, G. (2011). Cognitive load theory and multimedia learning, task characteristics and learning engagement: The Current State of the Art. *Computers in Human Behavior*, 27(1), 1–4.
- Klimsa, P. (Hrsg.). (2011). *Online-Lernen Handbuch für Wissenschaft und Praxis*. München: Oldenbourg.
- Koch, D. A. (1993). Testing goes graphical. *Journal of Interactive Instruction Development*, (5), 14–21.
- Koppel, I. (2011). Computerbasierte Förderdiagnostik: Usability-Anforderungen von Interfaces für funktionale Analphabet/inn/en. *bildungsforschung*, (2), 13–37.
- Koppel, I., Küster, J., & Wolf, K. D. (2013). Usability-Testing mit Funktionalen AnalphabetInnen. In A. Bereiter & C. Rensing (Hrsg.), *DeLFI 2013 – Die 11. e-Learning Fachtagung Informatik* (S. 293–296). Bonn: Gesellschaft für Informatik.
- Koppel, I., & Wolf, K. D. (2014). otu.lea: eine niedrigschwellige Online-Diagnostik für funktionale AnalphabetInnen in der Kursarbeit. *ALFA-FORUM*, 86.
- Kretschmann, R. (2006). Diagnostik bei Lernbehinderungen. In U. Petermann & F. Petermann (Hrsg.), *Diagnostik sonderpädagogischen Förderbedarfs* (S. 139–162). Göttingen: Hogrefe.
- Krug, S. (2002). *Don't make me think!* (2. Aufl.). Bonn: New Riders.
- Kuckartz, U., Rädiker, S., Ebert, T., & Schehl, J. (2013). *Statistik: Eine verständliche Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

-
- Kupka, K., Martens, A., & Diercks, J. (2011). Recruitment - wie Unternehmen auf spielerische Weise Bewerber gewinnen wollen. *Wirtschaftspsychologie aktuell*, 2, 53–56.
- Kyllonen, P. (2009). New constructs, methods and directions for computer-based assessment. In F. Sheuermann & J. Björnsson (Hrsg.), *he transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (S. 151–156). Luxemburg: Office for Official Publications of the European Communities.
- Laitusis, C. C. (2010). Examining the Impact of Audio Presentation on Tests of Reading Comprehension. *Applied Measurement in Education*, 23(2), 153–167.
- Laugwitz, B., Schrepp, M., & Held, T. (2006). Konstruktion eines Fragebogens zur Messung der User Experience von Softwareprodukten. *Proceedings of: Mensch & Computer 2006 - Mensch und Computer im Strukturwandel*, 125–134.
- Laugwitz, B., Schrepp, M., Held, T., Ilmberger, W., Tamm, N., & Schubert, U. (2009). Subjektive Benutzerzufriedenheit quantitativ erfassen: Erfahrungen mit dem User Experience Questionnaire UEQ. *Usability Professionals 2009*, 220–225.
- Leavitt, M. O., & Shneiderman, B. (2006). *Research-Based Web Design & Usability Guidelines* (Auflage: Rev Rep.). Washington, D.C.: U.S. Dept. of Health and Human Services.
- Legewie, H. (1994). In A. Boehm (Hrsg.), *Texte verstehen: Konzepte, Methoden, Werkzeuge* (S. 177–182). Konstanz: Univ.-Verl. Konstanz.
- Lenski, G., & Leggett, J. C. (1960). Caste, Class, and Deference. *American Journal of Sociology*, 65, 463–467.
- Lewis, C. (1982). Using the „thinking Aloud“ Method in Cognitive Interface Design. IBM T.J. Watson Research Center.

- Linde, A. (2007). Alphabetisierung, Grundbildung oder Literalität? In *Literalität, Grundbildung oder Lesekompetenz?: Beiträge zu einer Theorie-Praxis-Diskussion* (S. 90–99). Münster: Waxmann.
- Lindquist, E. F. (1950). Educational Measurement. (Washington: American Council for Education, Hrsg.).
- Low, R., & Jin, P. (2009). Motivation and Multimedia Learning. In R. Z. Zheng (Hrsg.), *Cognitive Effects of Multimedia Learning* (S. 154–172). Hershey: Idea Group Reference.
- Low, R., & Sweller, J. (2005). The Modality Principle in Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 147–167). New York: New York, NY, US: Cambridge University Press.
- Mason, B. J., Patry, M., & Bernstein, D. J. (2001). An examination of the equivalence between non-adaptive computer-based and traditional testing. *Journal of Educational Computing Research*, 24(1), 29–39.
- Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 31–48). New York: New York, NY, US: Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484–490.
- Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43–52.
- Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, 91(4), 638.
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of mul-

- timedia learning. *Journal of Educational Psychology*, 86(3), 389–401.
- Mayring, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung*, 1(2), 1–10.
- Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (Auflage: Neuauflage, 11., vollständig überarbeitete Aufl.). Weinheim: Beltz.
- Mayring, P., & Brunner, E. (2010). Qualitative Inhaltsanalyse. In B. Friebertshäuser, A. Langer, & A. Prengel (Hrsg.), *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft* (3. Aufl., S. 323–333). Weinheim [u.a.]: Juventa.
- McManu, S. (2008). Attributes of effective formative assessment. Washington D.C.: Council for Chief State School Officers.
- McNeill, K. L., & Krajcik, J. (2009). Synergy Between Teacher Practices and Curricular Scaffolds to Support Students in Using Domain-Specific and Domain-General Knowledge in Writing Arguments to Explain Phenomena. *Journal of the Learning Sciences*, 18(3), 416–460.
- Merriënboer, J. J. G. van, & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147–177.
- Middleton, J., Gorard, S., Taylor, C., & Bannan-Ritland, B. (2008). The „compleat“ design experiment: From soup to nuts. In A. E. Kelly, R. A. Lesh, & J. Y. Beak (Hrsg.), *Handbook of Design Research Methods in Education* (1. Aufl.). New York: Routledge.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, (63), 81–97.
- Molich, R., & Nielsen, J. (1990). Improving a human-computer dialogue. *Communications of ACM*, 33, 338–348.

- Montagnier, P. (2011). *Digital Divide: From Computer Access to Online Activities ? A Micro Data Analysis*. Paris: OECD Publishing.
- Moosbrugger, H. (2008). Item-Response-Theorie (IRT). In U.-P. D. H. Moosbrugger & D.-P. A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 215–259). Springer Berlin Heidelberg.
- Moray, N. (1979). *Mental workload: Its tehory and measurement*. New York: Plenum.
- Moreno, R. (2005). Multimedia Learning with Animated Pedagogical Agents. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 507–524). New York, NY, US: Cambridge University Press.
- Moreno, R., & Park, B. (2010). Cognitive Load Theory: Historical Development and RELation to Other Theories. In J. L. Plass, R. Moreno, & R. Brünken (Hrsg.), *Cognitive Load Theory* (S. 9–28). Cambridge University Press.
- Mosier, J. N., & Smith, S. L. (1986). Application of guidelines for designing user interface software. *Behaviour & Information Technology*, 5(1), 39–46.
- National Board of Medical Examiners (NBE). (2009). *Primum® Computer-based Case Simulations (CCS) for licensing doctors*.
- Neal, A. S., & Simons, R. M. (1984). Playback: a method for evaluating the usability of software and its documentation. *Proceedings of ACM CHI '83 Conference „Human Factors in Computing Systems“*, (78-82).
- Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C., & Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology*, 45(1), 21–27.
- Nickel, S. (2011). Literalitaet - Familie - family literacy die Transmission schriftkultureller Praxis und generationenuebergreifen-

- de Bildungsprogramme als Schluesselstrategie - Sowipor. *GESIS: Sowipor*, 35(3), 53–77.
- Nicol, D. J., & Milligan, C. (2006). Rethinking technology supported assessment in terms of the seven principles of good feedback practice. In C. Bryan & K. Clegg (Hrsg.), *Innovative Assessment in Higher Education* (S. 64–77). London: Routledge.
- Niegemann, H. M. (2008). *Kompndium multimediales Lernen*. Berlin; Heidelberg: Springer.
- Nielsen, J. (1992). Finding Usability problems through heuristic evaluation. *Proceedings of the ACM CHI '92 Conference*, 373–380.
- Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann.
- Nielsen, J. (1994). Usability Inspectin Methods. *Conference Companion CHI 1994*.
- Nielsen, J. (1997). The use and misuse of focus groups. *IEEE Software*, 14(1), 94–95.
- Nielsen, J. (2006). *Prioritizing Web usability*. Berkeley, Calif: New Riders.
- Nielsen, J. (2008). *Web Usability*. München: Addison-Wesley.
- Nielsen, J. (2010). Mental Models and User Experience Design. Abgerufen 23. September 2014, von <http://www.nngroup.com/articles/mental-models/>
- Nielsen, J. (2012). Usability 101: Introduction to Usability. Abgerufen 7. September 2014, von <http://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Nielsen, J. (2014). Severity Ratings for Usability Problems. Abgerufen 8. September 2014, von <http://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/>
- Nielsen, J., & Loranger, H. (2008). *Web Usability - Deutsche Ausgabe* (1. Aufl.). Addison-Wesley Verlag.

- Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods*. New York: Wiley John + Sons.
- Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceeding CHI '90*, 249–256.
- Nienkämper, B., & Bonna, F. (2010). Pädagogische Förderdiagnostik in der Grundbildung. *Der pädagogische Blick*, 18(4), 212–220.
- Niesyto, H. (2009). Digitale Medien, soziale Benachteiligung und soziale Distinktion. *Medienpädagogik – Zeitschrift Theorie und Praxis der Medienbildung*, 1–19.
- Norman, D. A. (1996). Dinge des Alltags. Gutes Design und Psychologie für Gebrauchsgegenstände. Campus Verlag GmbH.
- Norman, D. A. (1998). *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution* (Auflage: Revised.). Cambridge, Mass.: Mit Pr.
- Norris, N. (1997). Error, bias and validity in qualitative research. *Educational Action Research*, 5(1), 172–176.
- OECD. (1995). Grundqualifikationen, Wirtschaft und Gesellschaft. Ergebnisse der ersten internationalen Untersuchung von Grundqualifikationen Erwachsener.
- OECD (Hrsg.). (2000). Literacy in the Infomration Age.
- OECD (Hrsg.). (2005). *Learnig a Living. First Results of the Adult Literacy and Life Skills Survey*. Paris, Ottawa: OECD Publishing.
- Ossner, J. (2014). Rahmencurirriculum Schreiben. (Deutscher Volkshochschulverband e.V., Hrsg.).
- Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), 63–71.

-
- Pachler, N., Mellar, H., Daly, C., Yishay, M., & Wiliam, D. (2009). Scoping a vision for formative e-assessment: a project report for JISC. Institute of Education - University of London.
- Paivio, A. (1971). *Imagery and verbal processes*. New York u. a.: Holt, Rinehart and Winston.
- Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative Item Types for Computerized Testing. In W. J. van der Linden & G. A. W. Glas (Hrsg.), *Computerized Adaptive Testing: Theory and Practice* (S. 129–148). New York u. a.: Kluwer Academic Publishers.
- Parshall, C. G., & Harmes, J. C. (2007). Designing templates based on a taxonomy of innovaative items. *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. (National Research Council, Hrsg.). Washington: National Academies Press.
- Petermann, U., & Petermann, F. (2006). Zum Stellenwert sonderpädagogischer Förderdiagnostik. In U. Petermann & F. Petermann (Hrsg.), *Diagnostik sonderpädagogischen Förderbedarfs* (S. 1–15). Göttingen: Hogrefe.
- Petrie, H., & Bevan, N. (2009). The Evaluation of Accessibility, Usability and User Experience. In C. Stapanidis (Hrsg.), *The Universal Access Handbook*. Boca Raton: Taylor & Francis Group.
- Phye, G. D. (1996). *Handbook of Classroom Assessment: Learning, Achievement, and Adjustment*. Academic Press.
- Piaget, J., Aebli, H., & Montada, L. (2003). *Nachahmung, Spiel und Traum: die Entwicklung der Symbolfunktion beim Kinde* (5. Aufl.). Stuttgart: Klett-Cotta.
- Plass, J. L., Kalyuga, S., & Leutner, D. (2010). Individual Differences and Cognitive Load Theory. In J. L. Plass, R. Moreno, & R.

- Brünken (Hrsg.), *Cognitive Load Theory*. Cambridge University Press.
- Plass, J. L., Moreno, R., & Brünken, R. (Hrsg.). (2010). *Cognitive Load Theory*. Cambridge University Press.
- Pomplun, M., & Custer, M. (2005). The Score Comparability of Computerized and Paper-and-Pencil Formats for K-3 Reading Tests. *Journal of Educational Computing Research*, 32(2), 153–166.
- Pomplun, M., Ritchie, T., & Custer, M. (2006). Factors in Paper-and-Pencil and Computer Reading Score Differences at the Primary Grades. *Educational Assessment*, 11(2), 127–143.
- Posch, P., & Zehetmeier, S. (2010). Aktionsforschung. In S. Maschke & L. Stecher (Hrsg.), *EEO - Enzyklopädie Erziehungswissenschaft Online*. Weinheim: Juventa.
- Preece, J. (1994). *Human Computer Interaction* (1 edition.). Wokingham, England ; Reading, Mass: Addison Wesley.
- Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, 42(2), 185–217.
- Quante-Brandt, E., & Jäger, A. (Hrsg.). (2013). *lea.-Lernmaterialien*. Münster: Waxmann.
- Quellmalz, E. S., & Pellegrino, J. W. (2009). Technology and Testing. *Science*, 323(5910), 75–79.
- Raes, A., Schellens, T., De Wever, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education*, 59(1), 82–94.
- Rammstedt, B., Ackermann, D., Helmschrott, S., Klaukien, A., Maehler, D. B., Martin, S., ... Zabal, A. (Hrsg.). (2013). *PIAAC 2012: Die wichtigsten Ergebnisse im Überblick*. Münster: Waxmann.

-
- Rampl, H. (2007). Definition Usability. Abgerufen 7. September 2014, von <http://www.handbuch-usability.de/begriffsdefinition.html>
- Rauschenberger, M., Thomaschewski, T., & Schrepp, M. (2013). User Experience mit Fragebögen messen - Durchführung und Auswertung am Beispiel des UEQ. *DeLFI 2013-Die 11. e-Learning Fachtagung Infomratik*, 72–76.
- Reed, P., Holdaway, K., Isensee, S., Buie, E., Fox, J., Williams, J., & Lund, A. (1999). User interface guidelines and standards: progress, issues, and prospects. *Interacting with Computers*, 12(2), 119–142.
- Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist*, 41(2), 87–98.
- Reeves, T. (2006). Design Research from a technology perspective. In J. V. den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Hrsg.), *Educational Design Research* (S. 86–109). Routledge.
- Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den design-Based Research-Ansatz in der Lehr-Lernforschung. *Unterrichtswissenschaft*, 33(1), 52–69.
- Reuter-Liehr, C. (2008). Eine Einführung in das Training der phonemischen Strategie auf der Basis des rhythmischen Syllabierens mit einer Darstellung des Übergangs zur morphemischen Strategie. Bochum: Winkler.
- Rheinberg, F., & Krug, J. S. (2005). Motivationsförderung im Schulalltag: psychologische Grundlagen und praktische Durchführung (3., korrigierte Aufl.). Göttingen u.a.: Hogrefe.
- Richey, R. C., & Klein, J. D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education*, 16(2), 23–38.
- Ripley, M., Tafler, J., Ridgway, J., Harding, R., & Hakan, R. (2009). Review of Advanced e-Assessment Techniques. Abgerufen 6.

- November 2013, von <http://www.jisc.ac.uk/publications/reports/2009/raeatfinalreport.aspx>
- Rittmeyer, C. (2005). Kompendium Förderdiagnostik: Prinzipien, Methoden und Einsatzbereiche. Horneburg/Niederelbe: Persen.
- Rogers, Y., Preece, J., & Sharp, H. (2011). *Interaction design*. Hoboken, N.J.; Chichester: Wiley ; John Wiley [distributor].
- Rosebrock, C., Gausche, S., Haase, A., & Zimmer, D. (2014). Rahmencurriculum Lesen. (Deutscher Volkshochschulverband e.V., Hrsg.).
- Rosenblatt, B., & Lehmann, R. H. (2013a). Begrenzte Lernerfolge in Alphabetisierungskursen. *DIE aktuell*.
- Rosenblatt, B. von, & Lehmann, P. D. R. H. (2013b). Grade der Schriftbeherrschung und subjektiver Lernerfolg bei Teilnehmenden an Alphabetisierungskursen. *Zeitschrift für Erziehungswissenschaft*, 16(1), 55–77.
- Ruedel, C., & Mandel, S. (2010). E-Assessment: Einsatzszenarien und Erfahrungen an Hochschulen. Waxmann.
- Sari, E., & Herrington, J. (2013). Using Design-based research to investigate the design and development of an online community of practice for teacher professional development. *EdMedia 2013 World Conference on Educational Media and Technology*, 24–27.
- Sarodnick, F., & Brau, H. (2011). Methoden der Usability Evaluation: Wissenschaftliche Grundlagen und praktische Anwendung (2. Auflage.). Huber, Bern.
- Scalise, K., & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing „Intermediate Constraint“ Questions and Tasks for Technology Platforms. *The Journal of Technology, Learning and Assessment*, 4(6).

-
- Scalise, K., & Gifford, B. (2008). *Innovative item types: Intermediate constraint questions and tasks for computer-based testing*. gehalten auf der Annual meeting of the National Council on Measurement in Education, New York.
- Schäffer, B. (2006). Die Bildung Älterer mit neuen Medien: Zwischen Medienkompetenz, ICT-Literacy und generationsspezifischen Medienpraxiskulturen. *bildungsforschung.org*, 3(2), 1–28.
- Schladebach, A. (2007). Ein rotes Tuch: Formulare und Fragebögen! Auswertung der Teilnehmerbefragung im 2. Semester 2004 im Grundbildungszentrum der Hamburger Volkshochschule. In A. Grotlüschen & A. Linde (Hrsg.), *Literalität, Grundbildung oder Lesekompetenz? Beiträge zu einer Theorie-Praxis-Diskussion* (S. 140–146). Münster, New York, München u.a.: Waxmann.
- Schleicher, A. (2008). Piacac: A New Strategy for Assessing Adult Competencies. *International Review of Education*, 54(5-6), 627–650.
- Schnell, R., Hill, P. B., & Esser. (2011). *Methoden der empirischen Sozialforschung*. München: Oldenbourg, R.
- Schnotz, W. (2006). *Pädagogische Psychologie: Workbook*. Weinheim: Beltz, PVU.
- Schönrade, S., & Pütz, G. (2004). Die Abenteuer der kleinen Hexe: Bewegung und Wahrnehmung beobachten, verstehen, beurteilen, fördern (Auflage: 4.). Dortmund: borgmann publishing.
- Schwabe, G. (2001). Theorien zur Mediennutzung bei der Gruppenarbeit. In G. Schwabe, N. Streitz, & R. Unland (Hrsg.), *CSCW-Kompendium, Lehr- und Handbuch zum computergestützten kooperativen Arbeiten* (S. 54–65). Heidelberg: Springer.
- Schwartz, D. L., Chang, J., & Martin, L. (2008). In A. E. Kelly, R. A. Lesh, & J. Y. Beak (Hrsg.), *Handbook of Design Research Methods in Education* (1. Aufl., S. 47–67). New York: Routledge.

- Seale, C. (1999). *The Quality of Qualitative Research*. London ; Thousand Oaks, Calif: SAGE Publications Ltd.
- Shackel, B. (1985). Human factors and usability - whence and whither. In H.-J. Bullinger (Hrsg.), *Software-Ergonomie '85* (S. 13–31). Stuttgart: Teubner.
- Shackel, B. (1991). Usability - context, framework, design and evaluation. In B. Shackel & S. J. Richardson (Hrsg.), *Human Factors for Informatics Usability* (S. 21–38). Cambridge: Cambridge University Press.
- Sharit, J., Fisk, A. D., Rogers, W. A., Charness, N., & Czaja, S. J. (2007). *Designing for Older Adults* (1. Aufl.). Taylor & Francis.
- Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments*, 15(1), 27–46.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the Science of Education Design Studies. *Educational Researcher*, 32(1), 25–28.
- Shneiderman, B. (2000). Universal Usability. *Commun. ACM*, 43(5), 84–91.
- Shneiderman, B., & Plaisant, C. (2005). *Designing the User Interface*. Boston et al.: Pearson Education.
- Snow, R. E. (1993). Construct Validity and Constructed-Response Tests. In R. E. Bennett & W. C. Ward (Hrsg.), *Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment* (S. 45–60). Hillsdale, New Jersey: Lawrence Erlbaum.
- Snyder, C. (2007). Paper prototyping: the fast and easy way to design and refine user interfaces. Amsterdam u. a.: Kaufmann.
- Spieß, E. (1994). Aktionsforschung. In L. von Rosenstiel & W. Molt (Hrsg.), *Handbuch für angewandte Psychologie* (S. 1–8). Stuttgart: Poeschel.

-
- Spitta, G. (1997). Kinder schreiben eigene Texte: Klasse 1 und 2: Lesen und Schreiben im Zusammenhang ; spontanes Schreiben ; Schreibprojekte. Berlin: Cornelsen Scriptor.
- Squire, K. D. (2005). Resuscitating Research in Educational Technology: Using game-based learning research as a lens for looking at design-based research. *Educational Technology*, 45(1), 8–14.
- Statistisches Bundesamt. (2012). Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien. 4, 15, 1–34.
- Steinke, I. (2005). Gütekriterien qualitativer Forschung. In U. Flick, E. von Kardorff, & I. Steinke (Hrsg.), *Qualitative Forschung: Ein Handbuch* (Auflage: 10., S. 319–331). Reinbek bei Hamburg: Rowohlt.
- Steuten, U. (2013). Gesellschaftliche Teilhabe und funktionaler Alphabetismus. *ALFA-FORUM*, (84), 7–10.
- Sting, S. (2005). Literacy versus Schriftlichkeit. In J. Ecarius & B. Friebertshäuser (Hrsg.), *Literalität, Bildung und Biographie* (S. 18–38). Opladen: Barbara Budrich.
- Stöckberg, U. (2012). A research review of e-assessment. *Assessment & Evaluation in Higher Education*, 37(5), 591–604.
- Streeter, L., Bernstein, J., Foltz, P., & Donald, D. (2011). Pearson's Automated Scoring of Writing, Speaking, and Mathematics. Pearson.
- Sukkarieh, J. Z., & Stoyanchev, S. (2009). *Automating Model Building in c-rater*.
- Svinicki, M. (1999). New directions in learning. *New Directions for Teaching and Learning*, (80), 5–27.
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12, 29–46.

- Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81, 457–466.
- Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Hrsg.), *The Cambridge Handbook of Multimedia Learning* (S. 19–30). New York, NY, US: Cambridge University Press.
- Sweller, J. (2010). Cognitive Load Theory: Recent Theoretical Advances. In J. L. Plass, R. Moreno, & R. Brünken (Hrsg.), *Cognitive Load Theory* (S. 29–47). Cambridge University Press.
- Tabak, I. (2004). Synergy: A Complement to Emerging Patterns of Distributed Scaffolding. *Journal of the Learning Sciences*, 13(3), 305–335.
- Taras, M. (2005). Assessment – Summative and Formative – Some Theoretical Reflections. *British Journal of Educational Studies*, 53(4), 466–478.
- Taylor, S. J., & Bogdan, R. (1984). Introduction to qualitative research methods: the search for meanings. Wiley.
- The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32(1), 5–8.
- UNESCO. (1978). Revised Recommendation concerning the International Standardization of Educational Statistics. Abgerufen 24. Oktober 2014, von http://portal.unesco.org/en/ev.php-URL_ID=13136&URL_DO=DO_TOPIC&URL_SECTION=201.html
- United States Dept of Health and Human Services. (2006). *Research-based web design & usability guidelines*. U.S. Dept. of Health and Human Services.
- Von Rosenblatt, B., & Bilger, F. (2011). Erwachsene in Alphabetisierungskursen der Volkshochschulen - Ergebnisse einer reprä-

- sentativen Befragung (AlphaPanel). (Deutscher Volkshochschulverband, Hrsg.). Bonn.
- Walker, D. (2006). Toward Productive Design Studies. In J. V. den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Hrsg.), *Educational Design Research* (S. 9–44). Routledge.
- Wang, F., & Hannafin, M. J. (2005). Design-Based Research and Technology-Enhanced Learning Environments. *Educational Technology Research and Development*, 53(4), 5–23.
- Weidemann, B. (2002). Multicodierung und Multimodalität im Lernprozess. In *Informationen und Lernen mit Multimedia und Internet* (S. 45–62). Weinheim: Beltz.
- Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie*, 61(1), 161–265.
- Wilhelm, T., Tobias, V., Waltner, C., Hopf, M., & Wiesner, H. (2012). Design-Based Research am Beispiel der zweidimensional-dynamischen Mechanik. In *Konzepte fachdidaktischer Strukturierung* (Bd. 32, S. 31–47). Münster.
- Wirth, J., & Klieme, E. (2003). Computer-based Assessment of Problem Solving Competence. *Assessment in Education: Principles, Policy & Practice*, 10(3), 329–345.
- Wolf, K. D., & Koppel, I. (2010). Paper-Prototyping im Rahmen der Entwicklung von Instrumenten der Förderdiagnostik funktionaler Analphabeten. *Der Pädagogische Blick*, 18(4), 221–230.
- Wolf, K. D., & Koppel, I. (2014). Pädagogische Online-Diagnostik für funktionale Analphabeten - Zugänge öffnen und individualisiert fördern. *Argumente und Materialien der Hanns-Seidel-Stiftung*, (94), 9–19.
- Wolf, K. D., Koppel, I., & Schwedes, K. (2011). Potenziale von Rich E-Assessment für die Förderdiagnostik funktionaler Analphabeten. In A. Grotlüschen, R. Kretschmann, E. Quante-Brandt, & K. D. Wolf (Hrsg.), *Literalitätsentwicklung von funktionalen An-*

- alphabeten – Diagnose und Förderung auf den unteren Sprossen der Leiter* (S. 122–153). Münster: Waxmann.
- Wottawa, H. (2006). Evaluation. In A. Krapp & B. Weidemann (Hrsg.), *Pädagogische Psychologie: ein Lehrbuch* (S. 647–674). Weinheim [u.a.]: Beltz.
- Xie, B., & Salvendy, G. (2000). Prediction of Mental Workload in Single and Multiple Tasks Environments. *International Journal of Cognitive Ergonomics*, 4(3), 213–242.
- Zabal, A., Martin, S., Klaukien, A., Rammstedt, B., Baumert, J., & Klie-me, E. (2013). Grundlegende Kompetenzen der erwachsenen Bevölkerung in Deutschland im internationalen Vergleich. In B. Rammstedt (Hrsg.), *PIAAC 2012: Die wichtigsten Ergebnisse im Überblick* (S. 31–76). Münster: Waxmann.
- Zimmer. (2004). In C. Becker-Carus, F. Dorsch, & H. O. Häcker (Hrsg.), *Dorsch Psychologisches Wörterbuch: [15000 Stichwörter, 800 Testnachweise]*. Bern [u.a.]: Huber.

Webseiten

<http://blogs.epb.uni-hamburg.de/lea/>
(zuletzt geprüft am 03.11.2015)

www.bangor.ac.uk/cs/news/creating-a-computer-simulated-virtual-patient-to-train-clinicians-15617 (zuletzt geprüft am 03.11.2015)

www.de.cito.com
(zuletzt geprüft am 03.11.2015)

www.gcflearnfree.org
(zuletzt geprüft am 03.11.2015)

www.gov.uk/government/organisations/departement-for-business-innovation-skills (zuletzt geprüft am 03.11.2015)

www.grundbildung.de/daten/Grundlagen/definition
(zuletzt geprüft am 03.11.2015)

www.ich-will-lernen.de
(zuletzt geprüft am 03.11.2015)

www.latex-project.org
(zuletzt geprüft am 03.11.2015)

www.lernspiel-winterfest.de
(zuletzt geprüft am 03.11.2015)

www.medizinische-fakultaet-hd.uni-heidelberg.de/
(zuletzt geprüft am 03.11.2015)

www.nngroup.com/articles/usability-101-introduction-to-usability/
(zuletzt geprüft am 03.11.2015)

www.otulea.uni-bremen.de (zuletzt geprüft am 03.11.2015)

www.pearson.com.au/educator/secondary/digital-learning/pearson-reader-20/
(zuletzt geprüft am 03.11.2015)

www.r-project.org
(zuletzt geprüft am 03.11.2015)

www.riversity.activeworlds.com
(zuletzt geprüft am 03.11.2015)

www.stat.uni-muenchen.de/~leisch/Sweave
(zuletzt geprüft am 03.11.2015)

www.taotesting.com
(zuletzt geprüft am 03.11.2015)

www.targetskills.net
(zuletzt geprüft am 03.11.2015)

www.tinkering.net/vp
(zuletzt geprüft am 03.11.2015)

www.usmle.org/practice-materials/index.html
(zuletzt geprüft am 03.11.2015)

www.versanttest.com
(zuletzt geprüft am 03.11.2015)

Anhang

Im Anhang sind die für das Verständnis, die Transparenz sowie die Nachvollziehbarkeit relevanten Informationen aufgeführt. Weitere Dokumente und Daten (z.B. Auswertungsleitfäden, Zusammenfassung der Usability-Heuristiken) sind bei der Autorin unter ikoppel@uni-bremen.de auf Anfrage erhältlich.

Inhaltsverzeichnis

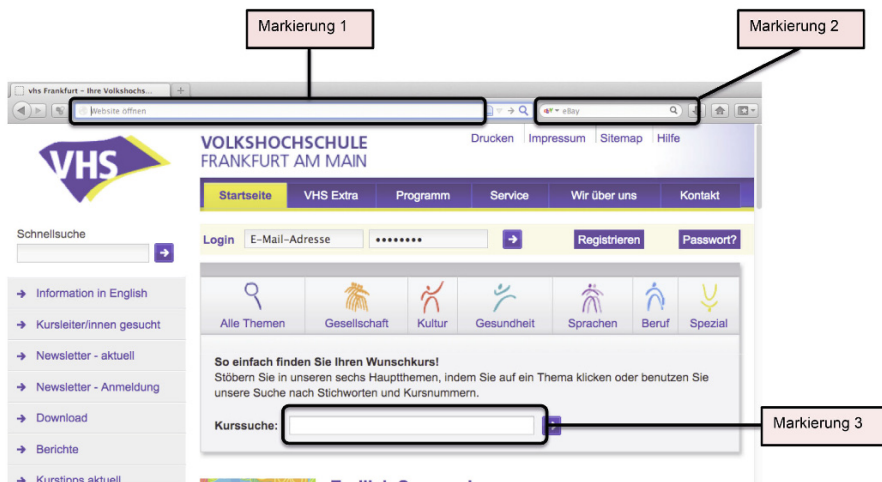
Tabellenverzeichnis	405
1. Erhebungsinstrumente	406
1.1 Fragebogen Computerkompetenz	406
1.2 Fragebogen zur Computererfahrung.....	408
1.3 User Experience Questionnaire	409
1.4 Leitfaden Fokussiertes Interview	411
1.5 Interviewleitfaden Experten-/ Expertinnengespräch....	413
2. Daten	414
2.1 Computerkompetenz.....	414
2.2 User Experience Questionnaire	415
2.3 Data-Logging.....	418

Tabellenverzeichnis

<i>Tabelle 1: Computerkompetenz - Eingabe Internetadresse.....</i>	<i>414</i>
<i>Tabelle 2: Computerkompetenz - Eingabe von Buchstaben in ein Textfeld....</i>	<i>414</i>
<i>Tabelle 3: Computerkompetenz - Eingabe Großschreibung</i>	<i>414</i>
<i>Tabelle 4: UEQ Ergebnisse.....</i>	<i>416</i>
<i>Tabelle 5: UEQ Konfidenzintervalle der Skalen.....</i>	<i>416</i>
<i>Tabelle 6: UEQ Mittelwerte der Skalen.....</i>	<i>416</i>
<i>Tabelle 7: UEQ Item-Korrelation der Skalen.....</i>	<i>417</i>
<i>Tabelle 8: Gesamtbearbeitungszeit.....</i>	<i>418</i>
<i>Tabelle 9: Deskriptive Darstellung Bearbeitungszeit.....</i>	<i>419</i>

1 Erhebungsinstrumente

1.1 Fragebogen Computerkompetenz



Frage 2: Sie möchten in ein Textfeld ein Wort eintippen. Wie gehen Sie vor?



+



☐ Ich klicke mit der markierten linken Maustaste in das Textfeld und tippe das Wort ein.



+



☐ Ich klicke mit der markierten rechten Maustaste in das Textfeld und tippe das Wort ein.



+



☐ Ich klicke mit der markierten mittleren Maustaste in das Textfeld und tippe das Wort ein.

?

☐ Weiß nicht

Frage 3: Sie möchten in ein Textfeld das Wort „Auto“ eingeben. Welche Tasten drücken Sie?



+


☐


+


☐


+


☐

?

Weiß nicht

☐

1.2 Fragebogen zur Computererfahrung

Fragebogen zur Person

Code: _____

Zum Schluss möchten wir Sie bitten, die folgenden Fragen zur Person zu beantworten:

1. Geschlecht: männlich ☐ weiblich ☐

2. Alter _____ Jahre

3. Wie gerne arbeiten Sie mit dem Computer auf einer Skala von 1-5?

Nicht gerne (1) ☐ ☐ ☐ ☐ ☐ (5) sehr gerne

4. Besitzen Sie einen Computer?

Ja ☐ nein ☐

5. Wie viel Zeit in Stunden verbringen Sie durchschnittlich pro Woche mit dem Computer? _____ Stunden

6. Wofür nutzen Sie den Computer?

7. Haben Sie Zugang zum Internet? (wenn nein, weiter mit 8.) Ja ☐ nein ☐

7a) Wie viel Zeit verbringen Sie durchschnittlich pro Woche aktiv im Internet? _____ Stunden

7b) Für welche Zwecke nutzen Sie das Internet?	mind. 1x am Tag	mind. 1x Woche	weniger als 1x Woche	nie
1. Surfen	☐	☐	☐	☐
2. Informationensuche	☐	☐	☐	☐
3. Informationen über aktuelle Nachrichten?	☐	☐	☐	☐
4. Email	☐	☐	☐	☐
5. Einkaufen	☐	☐	☐	☐
6. Telefonieren mit dem Computer	☐	☐	☐	☐
7. Musik hören	☐	☐	☐	☐
8. Videos ansehen	☐	☐	☐	☐
9. Fernsehen	☐	☐	☐	☐
10. Computerspiele	☐	☐	☐	☐
11. Soziale Netzwerke (z.B. Facebook)	☐	☐	☐	☐
12. Sonstiges _____	☐	☐	☐	☐

Was möchten Sie uns noch mitteilen?

Nochmals vielen Dank für Ihre Mithilfe!

1.3 User Experience Questionnaire

Code _____

Bitte geben Sie Ihre Beurteilung ab (bitte die Gegensatzpaare vorlesen!).

Um das Produkt zu bewerten, möchten wir mit Ihnen diesen Fragebogen ausfüllen und Ihre persönliche Meinung erfahren. Er besteht aus Gegensatzpaaren von Eigenschaften, die das Produkt haben kann. Abstufungen zwischen den Gegensätzen sind durch Kreise dargestellt. Durch Ankreuzen eines dieser Kreise können Sie Ihre Zustimmung zu einem Begriff äußern. Ich werde nur das ankreuzen, was Sie mir vorgeben.

Beispiel:

attraktiv	      	unattraktiv

Mit dieser Beurteilung sagen Sie aus, dass Sie das Produkt eher attraktiv als unattraktiv einschätzen.

Entscheiden Sie möglichst spontan. Es ist wichtig, dass Sie nicht lange über die Begriffe nachdenken, damit Ihre unmittelbare Einschätzung zum Tragen kommt.

Bitte kreuzen Sie immer eine Antwort an, auch wenn Sie bei der Einschätzung zu einem Begriffspaar unsicher sind oder finden, dass es nicht so gut zum Produkt passt.

Es gibt keine „richtige“ oder „falsche“ Antwort. Ihre persönliche Meinung zählt!

Bitte geben Sie nun Ihre Einschätzung des Produkts ab. Kreuzen Sie bitte nur einen Kreis pro Zeile an.

Code _____

	1	2	3	4	5	6	7		
unerfreulich	☐	☐	☐	☐	☐	☐	☐	erfreulich	1
unverständlich	☐	☐	☐	☐	☐	☐	☐	verständlich	2
kreativ	☐	☐	☐	☐	☐	☐	☐	phantasielos	3
leicht zu lernen	☐	☐	☐	☐	☐	☐	☐	schwer zu lernen	4
erfrischend	☐	☐	☐	☐	☐	☐	☐	einschläfernd	5
langweilig	☐	☐	☐	☐	☐	☐	☐	spannend	6
uninteressant	☐	☐	☐	☐	☐	☐	☐	interessant	7
unberechenbar	☐	☐	☐	☐	☐	☐	☐	voraussagbar	8
schnell	☐	☐	☐	☐	☐	☐	☐	langsam	9
neu	☐	☐	☐	☐	☐	☐	☐	alt	10
unbedienbar	☐	☐	☐	☐	☐	☐	☐	bedienbar	11
gut	☐	☐	☐	☐	☐	☐	☐	schlecht	12
kompliziert	☐	☐	☐	☐	☐	☐	☐	einfach	13
abstoßend	☐	☐	☐	☐	☐	☐	☐	anziehend	14
veraltet	☐	☐	☐	☐	☐	☐	☐	modern	15
unangenehm	☐	☐	☐	☐	☐	☐	☐	angenehm	16
vorhersagbar	☐	☐	☐	☐	☐	☐	☐	unvorhersagbar	17
abwechslungsreich	☐	☐	☐	☐	☐	☐	☐	eintönig	18
zuverlässig	☐	☐	☐	☐	☐	☐	☐	unzuverlässig	19
(ineffizient)	☐	☐	☐	☐	☐	☐	☐	(effizient)	20
übersichtlich	☐	☐	☐	☐	☐	☐	☐	verwirrend	21
stockend	☐	☐	☐	☐	☐	☐	☐	flüssig	22
aufgeräumt	☐	☐	☐	☐	☐	☐	☐	überladen	23
schön	☐	☐	☐	☐	☐	☐	☐	hässlich	24
sympathisch	☐	☐	☐	☐	☐	☐	☐	unsympathisch	25
unauffällig	☐	☐	☐	☐	☐	☐	☐	auffällig	26

(AutorInnen: Laugwitz, B./ Schrepp, M. / Held, T.; abrufbar unter <http://www.ueq-online.org/>)

1.4 Leitfaden Fokussiertes Interview

Code: _____

Einleitung: Ich stelle Ihnen nun ein paar Fragen zu dem Test. Ich möchte Ihre persönliche Meinung erfahren und wissen, was Sie gemacht haben. Sind Sie damit einverstanden, dass wir die Befragung aufnehmen? Ich werde keine Fragen stellen, die mit Ihrer Person in Verbindung zu bringen sind. Die Befragung ist anonym. Es gibt nicht richtig oder falsch und es geht nicht darum, Sie zu testen.

Einstiegsfragen:

1. Wie war es für Sie, das Programm zu nutzen?
2. Welche besonderen Erfahrungen haben Sie gemacht?
 - Was hat Ihnen besonders Spaß gemacht?
 - Gab es Schwierigkeiten/ Hatten Sie Probleme? (wenn nein, weiter mit 3.)
 - Falls ja, konnten Sie diese lösen?

Erlernbarkeit:

3. Wie war der Anfang bzw. der Einstieg für Sie?
4. Welche Erfahrungen haben Sie mit der Hilfefunktion gemacht?

Kognitive Belastung:

5. Wie konnten Sie die gesprochene Anweisungen verstehen?
 - Ton/Inhalt?
6. Was fällt Ihnen zum Aufbau/ zur Struktur des Programms ein?
7. Erzählen Sie bitte, wie Sie sich zurechtgefunden haben.
 - Gab es Hinweise, die Ihnen den Umgang mit dem Programm erleichtert haben?
8. Wünschen Sie sich mehr Informationen, wie das Programm zu nutzen ist?
 - Wenn ja, an welchen Stellen?

Motivation:

9. Wie motiviert waren Sie, das Programm zu nutzen?
10. Was fällt Ihnen zu den Geschichten und den Figuren im Programm ein?
 - Konnten Sie sich in die Personen hineinversetzen?
11. Würden Sie gerne öfter mit diesem Programm arbeiten?

12. Fanden Sie die Aufgaben eher schwer oder eher leicht?

Design:

13. Wie gefällt Ihnen die farbliche Gestaltung?

14. Was ist Ihnen in Bezug auf die Schrift aufgefallen?

- Ggf. Screenshot zeigen

Empfindung:

15. Wie fühlen Sie sich nach dem Test?

- Z.B. erledigt, zufrieden, erschöpft, motiviert?

Abschlussfrage: Nun habe ich Ihnen viele Fragen gestellt. Gibt es noch etwas, das Sie uns fragen möchten? Haben Sie Anmerkungen/ Lob/ Kritik?

Vielen Dank nochmals, dass Sie teilgenommen haben!

1.5 Interviewleitfaden Experten-/ Expertinnengespräch

Einleitung: Vielen Dank, dass Sie sich die Zeit nehmen und Sie bereit sind, ein Gespräch mit mir zu führen. Ist es in Ordnung, wenn ich das Gespräch aufzeichne? Es werden keine personenbezogenen Daten gespeichert.

1. Wie arbeiten Sie in den Alphabetisierungskursen?
2. Haben Sie Erfahrungen mit dem Einsatz von Computern in Alphabetisierungskursen?
3. Falls ja: Welche Programme nutzen Sie?
4. Wie ist das für Sie, Computer einzusetzen?
5. Hilfreich, unterstützend, entlastend, stressig, aufregend...?
6. Kennen Sie otu.lea?
7. Falls ja: Was fällt Ihnen zu otu.lea ein?
8. Würden Sie otu.lea einsetzen? Warum? Warum nicht?
9. Was brauchen Sie/die TN, damit Sie es gerne einsetzen?
10. Hätten Sie Lust, sich das Programm anzusehen, wenn es fertig entwickelt ist und uns eine Rückmeldung zu geben? (Falls ja: Kontaktdaten!)
11. Dürfte ich noch kurz erfahren, wie lange Sie bereits als Kursleitende in Alphabetisierungskursen tätig sind und wie viele Kurse Sie momentan betreuen?

Nochmals vielen Dank, dass Sie sich die Zeit genommen haben, dass Sie an diesem Gespräch teilgenommen haben und dass Sie Ihre Teilnehmenden hierher begleitet haben!

2 Daten

2.1 Computerkompetenz

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	richtig	45	90,0	93,8	93,8
	falsch	3	6,0	6,3	100,0
	Gesamtsumme	48	96,0	100,0	
Fehlend	System	2	4,0		
	Gesamtsumme	50	100,0		

Tabelle 1: Computerkompetenz - Eingabe Internetadresse

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	richtig	46	92,0	92,0	92,0
	falsch	3	6,0	6,0	98,0
	weiß nicht	1	2,0	2,0	100,0
	Gesamtsumme	50	100,0	100,0	

Tabelle 2: Computerkompetenz - Eingabe von Buchstaben in ein Textfeld

		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozente
Gültig	richtig	35	70,0	70,0	70,0
	falsch	15	30,0	30,0	100,0
	Gesamtsumme	50	100,0	100,0	

Tabelle 3: Computerkompetenz - Eingabe Großschreibung

2.2 User Experience Questionnaire

Item	MW	Varianz	SD	Anzahl	Positiv	Negativ	Skala
1	1,7	1,1	1,1	10	unerfreulich	erfreulich	Attraktivität
2	2,0	1,3	1,2	10	unverständlich	verständlich	Durchschaubarkeit
3	1,2	0,8	0,9	10	kreativ	phantasielos	Originalität
4	2,0	1,8	1,3	10	leicht zu lernen	schwer zu lernen	Durchschaubarkeit
5	1,6	0,9	1,0	10	erfrischend	einschläfernd	Stimulation
6	0,9	0,5	0,7	10	langweilig	spannend	Stimulation
7	1,3	1,6	1,3	10	uninteressant	interessant	Stimulation
8	1,3	0,9	0,9	10	unberechenbar	voraussagbar	Steuerbarkeit
9	1,4	1,4	1,2	10	schnell	langsam	Effizienz
10	0,5	1,2	1,1	10	neu	alt	Originalität
11	1,6	1,6	1,3	10	unbedienbar	bedienbar	Steuerbarkeit
12	2,4	0,5	0,7	10	gut	schlecht	Attraktivität
13	1,8	2,2	1,5	10	kompliziert	einfach	Durchschaubarkeit
14	1,5	0,7	0,8	10	abstoßend	anziehend	Attraktivität
15	1,0	0,8	0,9	9	veraltet	modern	Originalität
16	1,9	1,4	1,2	10	unangenehm	angenehm	Attraktivität
17	1,7	1,1	1,1	10	vorhersagbar	unvorhersagbar	Steuerbarkeit
18	1,1	0,8	0,9	10	abwechslungsreich	eintönig	Stimulation
19	1,7	0,9	0,9	10	zuverlässig	unzuverlässig	Steuerbarkeit
20	1,8	0,4	0,6	10	ineffizient	effizient	Effizienz
21	2,0	2,2	1,5	10	übersichtlich	verwirrend	Durchschau-

					lich		barkeit
22	1,7	0,9	0,9	10	stockend	flüssig	Effizienz
23	1,8	0,8	0,9	10	aufge- räumt	überladen	Effizienz
24	1,6	1,6	1,3	10	schön	hässlich	Attraktivität
25	2,0	0,7	0,8	10	sympa- thisch	unsympa- thisch	Attraktivität
26	0,6	1,6	1,3	10	unauffällig	auffällig	Originalität

Tabelle 4: UEQ Ergebnisse

Konfidenzintervalle (p=0.05) pro Skala						
Skala	MW	SD	N	Konfidenz	Konfidenzintervall	
Attraktivität	1,850	0,791	10	0,490	1,360	2,340
Durchschaubarkeit	1,950	1,026	10	0,636	1,314	2,586
Effizienz	1,675	0,624	10	0,387	1,288	2,062
Steuerbarkeit	1,575	0,736	10	0,456	1,119	2,031
Stimulation	1,225	0,854	10	0,529	0,696	1,754
Originalität	0,825	0,635	10	0,394	0,431	1,219

Tabelle 5: UEQ Konfidenzintervalle der Skalen

Skalen	MW
Attraktivität	1,850
Durchschaubarkeit	1,950
Effizienz	1,675
Steuerbarkeit	1,575
Stimulation	1,225
Originalität	0,825

Tabelle 6: UEQ Mittelwerte der Skalen

Attraktivität	
Items	Korrelation
1, 12	0,63
1, 14	0,68
1, 16	0,67
1, 24	0,65
1, 25	0,51
12, 14	0,19
12, 16	0,85
12, 24	0,45
12, 25	0,58
14, 16	0,16
14, 24	0,72
14, 25	0,64
16, 24	0,48
16, 2	0,45
24, 25	0,86
DK	0,57
Alpha	0,89

Durchschaubarkeit	
Items	Korrelation
2, 4	0,87
2, 13	0,85
2, 21	-0,06
4, 13	0,90
4, 21	-0,06
13, 21	0,15
DK	0,44
Alpha	0,76

Effizienz	
Items	Korrelation
9, 20	0,87
9, 22	0,02
9, 23	0,08
20, 22	0,26
20, 23	0,11
22, 23	0,43
DK	0,30
Alpha	0,63

Steuerbarkeit	
Items	Korrelation
8, 11	0,57
8, 17	0,21
8, 19	0,48
11, 17	-0,02
11, 19	0,35
17, 19	0,34
DK	0,32
Alpha	0,66

Stimulation	
Items	Korrelation
5, 6	0,56
5, 7	0,94
5, 18	0,84
6, 7	0,64
6, 18	0,53
7, 18	0,68
DK	0,70
Alpha	0,90

Originalität	
Items	Korrelation
3, 10	0,22
3, 15	0,89
3, 26	-0,11
10, 15	0,38
10, 26	0,00
15, 26	0,00
DK	0,23
Alpha	0,55

Tabelle 7: UEQ Item-Korrelation der Skalen

2.3 Data-Logging

Person	Gesamtbearbeitungszeit
1OZDE	00:47:59
2XCDB	00:34:20
2YCEU	00:34:56
34CYK	00:30:48
3IC2Y	00:28:45
3YZDE	00:23:05
4ORAE	00:33:56
4X8G2	00:26:00
5LC8B	00:38:25
67OHY	00:20:20
6CYZA	00:23:12
6SCDK	00:45:04
7K52E	00:59:40
84A5X	00:58:58
857FR	00:31:20
8BFTZ	01:12:10
8TRD1	00:46:05
937D2	00:23:50
9BSD5	00:42:30
9CLF5	00:30:35
ALDET	00:52:12
EAM5U	00:23:45
EWL9D	00:38:45
FMD48	00:21:24
G8AB6	00:26:50

Person	Gesamtbearbeitungszeit
HB754	00:32:22
HYLDZ	00:17:24
ICGZW	00:26:56
K5C3U	00:56:26
L453K	00:42:47
L8XDF	00:25:05
MALXD	00:52:46
OLMB7	00:32:10
RAFXZ	00:20:45
RGCA7	00:28:10
RTL DG	00:25:40
SBIDE	00:40:41
SXMDE	00:51:34
TOYDU	00:29:34
WBODR	01:22:18
X9E23	00:46:53
XBI9G	00:33:27
XGFLM	00:18:12
YA8X1	00:39:28
YAIHE	00:33:00
YG5SE	00:41:50
YS3RL	00:20:45
Z27D3	00:11:00
Z76IA	00:42:03

Tabelle 8: Gesamtbearbeitungszeit

	N	Minimum	Maximum	MW	SD
Gesamtbearbeitungszeit	49	0:11:00	1:22:18	0:36:02	0:14:26

Tabelle 9: Deskriptive Darstellung Bearbeitungszeit