



UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Zweitschriftlernende in der deutschsprachigen Schweiz

In 300 Kurslektionen Alpha und A1?

Dissertation zur Erlangung der Doktorwürde an der Philosophischen Fakultät der Universität Freiburg (Schweiz).

Santi Guerrero Calle, Heimenhausen (BE), 2020

Genehmigt von der Philosophischen Fakultät auf Antrag von

Herrn Prof. Dr. Thomas Studer (1. Gutachter)

Frau Prof. Dr. Karen Schramm (2. Gutachterin)

Herrn Dr. Alexis Feldmeier García (3. Gutachter)

Freiburg, den 05.06.2020. Prof. Bernadette Charlier Pasquier, Dekanin.

Inhaltsverzeichnis

ABKÜRZUNGSVERZEICHNIS	V
ABBILDUNGSVERZEICHNIS	VI
TABELLENVERZEICHNIS	VII
EINLEITUNG.....	1
TEIL I: GRUNDLAGEN	8
1. KONZEPTUELLE UND EMPIRISCHE GRUNDLAGEN	8
1.1 SCHLÜSSELBEGRIFFE	8
1.1.1 <i>Flüchtlinge</i>	8
1.1.2 <i>Erstsprache und Erstschriftsprache</i>	9
1.1.3 <i>Analphabetismus, Grundbildung und Literalität</i>	10
1.2 ZWEITSCHRIFTERWERB	12
1.2.1 <i>Schriftspracherwerb</i>	12
1.2.2 <i>Transfer</i>	15
1.2.3 <i>Forschungsstand Zweitschrifterwerb</i>	16
1.3 ZWEITSCHRIFTLERNENDE	18
1.3.1 <i>Forschungsstand Zweitschriftlernende</i>	21
1.3.2 <i>Die Rolle des Bildungshintergrunds</i>	26
1.4 DIE SILBE – EIN INNOVATIVES LEHR-/LERNKONZEPT?	29
1.4.1 <i>Die Silbenmethode im Unterricht</i>	30
1.4.2 <i>Phonologische Bewusstheit</i>	33
1.4.3 <i>Forschungsstand zur Silbenmethode</i>	35
2. KURSKONZEPTE FÜR ZWEITSCHRIFTLERNENDE.....	36
2.1 ZWEITSCHRIFTLERNERKURSE DEUTSCHLAND	36
2.1.1 <i>Rahmenbedingungen Zweitschriftlernerkurs</i>	37
2.1.1.1 <i>Kursteilnehmer und Einstufung</i>	37
2.1.1.2 <i>Kursgefäß</i>	39
2.1.2 <i>Inhalte und Ziele des Basis-Sprachkurses</i>	40
2.1.2.1 <i>Erster Kursabschnitt</i>	40
2.1.2.2 <i>Zweiter und dritter Kursabschnitt</i>	41
2.1.3 <i>Didaktik im Basis-Sprachkurs</i>	42
2.1.4 <i>Qualifikation der Lehrpersonen</i>	42
2.2 KURSE IN DER DEUTSCHSCHWEIZ	43
2.2.1 <i>Kantonales Rahmenkonzept in Bern</i>	43
2.2.2 <i>Vergleich zu Deutschland</i>	45
TEIL II: METHODOLOGIE	47
3. FORSCHUNGSGEGENSTAND	47
3.1 EXPERIMENTAL- UND KONTROLLGRUPPE	49
3.2 ZWEITSCHRIFTLERNERKURSE	49
3.3 STICHPROBENBILDUNG UND AKTEURE	51
3.3.1 <i>Kurseinteilung</i>	53
3.3.2 <i>Sample</i>	54
3.3.3 <i>Lehrpersonen</i>	57
3.4 FORSCHUNGSETHIK BEIM TESTEN VON FLÜCHTLINGEN	58
3.5 FORSCHUNGSFRAGEN	61
3.6 HYPOTHESEN	61
4. TESTKONSTRUKTION	64
4.1 KOMMUNIKATIVE UND ALLGEMEINE SPRACHKOMPETENZ	64
4.2 TESTINSTRUMENTE	65

4.2.1	<i>Elicited Imitation Task</i>	68
4.2.1.1	Testkonstrukt	69
4.2.1.2	Aufgabenformat	71
4.2.1.3	Bewertungsverfahren.....	75
4.2.2	<i>Kunstwörter</i>	76
4.2.2.1	Testkonstrukt	78
4.2.2.2	Aufgabenformat	78
4.2.2.3	Bewertungsverfahren.....	79
4.2.3	<i>Oral Reading Fluency</i>	80
4.2.3.1	Testkonstrukt	82
4.2.3.2	Aufgabenformat	83
4.2.3.3	Bewertungsverfahren.....	84
4.2.4	<i>Hörverstehen</i>	85
4.2.4.1	Testkonstrukt	86
4.2.4.2	Aufgabenformat	87
4.2.4.3	Bewertungsverfahren.....	90
4.2.5	<i>Diktat</i>	91
4.2.5.1	Testkonstrukt	92
4.2.5.2	Aufgabenformat	92
4.2.5.3	Bewertungsverfahren.....	97
4.3	FRAGEBÖGEN	98
4.3.1	<i>Lehrerfragebogen</i>	98
4.3.2	<i>Lernerfragebogen</i>	99
5.	DATENGRUNDLAGE	102
5.1	PILOTIERUNG UND ÜBERARBEITUNG	102
5.1.1	<i>Erste Erprobung</i>	102
5.1.2	<i>Zweite Erprobung</i>	105
5.2	TESTABLAUF	105
5.3	KODIERUNG DER VARIABLEN	107
5.3.1	<i>Abhängige Variablen</i>	107
5.3.2	<i>Unabhängige Variablen</i>	107
5.3.2.1	Lehr- und Lernumgebungsvariablen.....	108
5.3.2.2	Unterrichtsbezogene Variablen.....	109
5.3.2.3	Personenbezogene Variablen.....	109
5.3.3	<i>Dummy-Variablen und Referenzkategorien</i>	112
5.3.4	<i>Stör- und Kontrollvariablen</i>	112
5.4	DATENAUFBEREITUNG	114
TEIL III:	ERGEBNISSE	115
6.	ANALYSE	115
6.1	TESTGÜTE	115
6.1.1	<i>Objektivität</i>	116
6.1.2	<i>Reliabilität</i>	117
6.1.2.1	Interne Konsistenz.....	117
6.1.2.2	Retest-Reliabilität	122
6.1.2.3	Bewerter-Reliabilität	126
6.1.3	<i>Validität</i>	127
6.1.3.1	Konstruktvalidität	127
6.1.3.2	Inhaltsvalidität	127
6.1.3.3	Kriteriumsvalidität	127
6.2	DESKRIPTIVE ANALYSE DER TELC-PRÜFUNGSRESULTATE	129
6.2.1	<i>Uni- und bivariate Analyse</i>	131
6.2.2	<i>Die Treatments im Vergleich</i>	141
6.2.3	<i>Regression</i>	143
6.2.4	<i>Überprüfung der Hypothesen</i>	146

6.3	DESKRIPTIVE ANALYSE DES GESAMTLERNERFOLGS BEI DEN FORSCHERGENERIERTEN TESTS	148
6.3.1	Uni- und bivariate Analyse	149
6.3.2	Die Treatments im Kursverlauf	155
6.3.3	Die Treatments bei den forschergenerierten Tests	157
6.3.4	Regression	160
6.3.5	Überprüfung der Hypothesen	162
6.4	ANALYSE DER KUNSTWÖRTER	163
6.4.1	Überprüfung der Hypothesen	164
TEIL IV: DISKUSSION		166
7.	DISKUSSION DER EMPIRISCHEN ERGEBNISSE	166
7.1	NIVEAU A1 INNERHALB VON 300 KURSLEKTIONEN (FORSCHUNGSFRAGE 1)	166
7.2	SILBENMETHODE (FORSCHUNGSFRAGE 2)	167
7.3	DISKUSSION DER ÜBRIGEN HYPOTHESEN	170
7.3.1	Schulerfahrung	171
7.3.2	Schreibkompetenz in der L1	172
7.3.3	Arbeitsgedächtnis	172
7.4	WEITERE ERGEBNISSE	174
8.	LIMITATIONEN UND GENERALISIERBARKEIT DER STUDIE	180
8.1	TESTINSTRUMENTE UND FORSCHUNGSDESIGN	180
8.2	FORSCHUNG MIT EXTERNEN PARTNERN	183
8.3	GENERALISIERBARKEIT	186
9.	IMPLIKATIONEN UND AUSBLICK	189
9.1	FORSCHUNG	189
9.2	BILDUNGSPOLITIK	190
9.3	PRAXIS	192
10.	SCHLUSSWORT	193
BIBLIOGRAFIE		195
ANHANG		213
1.	Beispiel eines Silbenprotokolls	213
2.	Unterrichtsbeispiele Silbenmethode	214
3.	Hörverstehen Einstufungstest	215
4.	Hörverstehen Follow-up 1	218
5.	Hörverstehen Follow-up 2	221
6.	Diktatvorlage	224
7.	Lehrerfragebogen	225
8.	Lernerfragebogen Deutsch	231
ANHANG, SEPARAT GEBUNDEN		236
	Inhaltsverzeichnis	236
ANHANG, CD-ROM		238
	Inhaltsverzeichnis	238

Danksagung

Ein umfangreiches Forschungsprojekt wie das vorliegende ist nicht ohne Unterstützung möglich, weshalb ich mich an dieser Stelle bei verschiedenen Personen und Institutionen herzlich bedanken möchte.

Zuerst möchte ich der Erziehungsdirektion des Kantons Bern für die Subventionierung der sechs Zweitschriftlernerkurse danken. Ohne diese Unterstützung wäre diese Studie nicht in diesem Umfang möglich gewesen. Ausserdem möchte ich den beiden Institutionen, ihrem Sekretariat und den Lehrpersonen für ihr Engagement, ihr Durchhaltevermögen und den fachlichen Austausch – der die Arbeit stets verbesserte – danken. Ein spezieller Dank geht an dieser Stelle an ECAP Bern, die als Kursinstitution die Weiterführung des Projekts ermöglichte.

Ein besonderer Dank gebührt meinem Betreuer und Mentor Prof. Dr. Thomas Studer, der mir über den ganzen Weg hinweg stets mit wertvollem fachlichen Rat zur Seite stand und an mich geglaubt hat. Für den fachlichen Austausch, die hilfreichen Anregungen und für die guten Zusprüche bedanke ich mich zudem bei den Teilnehmenden der Doktorandenkolloquien.

Von ganzem Herzen danke ich Keanu, Rajah, Rita und Yannick für ihre wertvolle Unterstützung bei der Datenerhebung. Rita und Yannick, ich danke euch dafür, dass ihr mir stets Mut zugesprochen habt und mir immer eine Stütze wart.

Ich möchte mich auch ganz herzlich bei meinen Übersetzern bedanken, die mir während der ganzen Phase mit wertvollen Inputs zur Seite gestanden sind.

Zu guter Letzt geht mein Dank an die Kursteilnehmenden, denn ohne diese wäre die Forschungsarbeit nicht möglich gewesen:

ལུགས་རྩེ་ཆེ་

நன்றி

تشکر

شکرا

የቀንየለይ።

Abkürzungsverzeichnis

ALTE	Association of Language Testers in Europe
AV	Abhängige Variable
BAMF	Bundesamt für Migration und Flüchtlinge
CBM	Curriculum-based measurement
CLB	Canadian Language Benchmarks
DaZ	Deutsch als Zweitsprache
EI	Elicited Imitation (Task)
ESL	English Second Language Learners
GER	Gemeinsamer europäischer Referenzrahmen für Sprachen
gGmbH	gemeinnützige GmbH im deutschen Steuerrecht
HV	Hörverstehen
KIP	Kantonales Integrationsprogramm
L1	Erstsprache
L2	Zweitsprache
L-B-Z	Laut-Buchstaben-Zuordnung
LESLLA	Literacy Education and Second Language Learning for Adults
LRS	Lese-Rechtschreibstörung
ORF	Oral Reading Fluency
telc	The European Language Certificates
TOEFL	Test for English as a Foreign Language
UV	Unabhängige Variable
WpM	Wörter pro Minute

Abbildungsverzeichnis

Abb. 1: HEIMATLÄNDER DER TEILNEHMENDEN	55
Abb. 2: GESCHLECHTERVERTEILUNG	56
Abb. 3: ANZAHL DER BESUCHTEN SCHULJAHRE IM HEIMATLAND	57
Abb. 4: PROZESSE BEI DER ELICITED IMITATION	70
Abb. 5: SELEKTIVES HÖRVERSTEHEN, EINSTUFUNGSTEST	88
Abb. 6: GLOBALES HÖRVERSTEHEN, EINSTUFUNGSTEST	89
Abb. 7: DETAILLIERTES HÖRVERSTEHEN, ZU ALLEN DREI MESSZEITPUNKTEN	90
Abb. 8: DIKTATVORLAGE	94
Abb. 9: GRUPPIERTE UNABHÄNGIGE VARIABLEN	108
Abb. 10: OBJEKTIVITÄT, RELIABILITÄT UND VALIDITÄT DER FORSCHERGENERIERTEN TESTS	115
Abb. 11: RELIABILITÄTSANALYSE ELICITED IMITATION	118
Abb. 12: RELIABILITÄTSANALYSE KUNSTWÖRTER	119
Abb. 13: RELIABILITÄTSANALYSE HÖRVERSTEHEN	120
Abb. 14: RELIABILITÄTSANALYSE DIKTAT	122
Abb. 15: TELC-PROZENTWERTE UND NORMALVERTEILUNG	129
Abb. 16: TELC PASS/FAIL	131
Abb. 17: TELC-PROZENTWERTE	132
Abb. 18: TELC PASS/FAIL DIFFERENZIERT NACH INSTITUTION	132
Abb. 19: TELC-PROZENTWERTE DIFFERENZIERT NACH INSTITUTION	133
Abb. 20: TELC-PROZENTWERTE DIFFERENZIERT NACH KURSLEVEL	134
Abb. 21: TELC-PROZENTWERTE DIFFERENZIERT NACH L1	135
Abb. 22: TELC-PROZENTWERTE DIFFERENZIERT NACH VORGÄNGIG BESUCHTEM DEUTSCHKURS	139
Abb. 23: TELC-PROZENTWERTE DIFFERENZIERT NACH KURSLEVEL UND SILBENMETHODE	140
Abb. 24: TELC-PROZENTWERTE HÖREN DIFFERENZIERT NACH KURSLEVEL UND SILBENMETHODE	143
Abb. 25: TELC-PROZENTWERTE SPRECHEN DIFFERENZIERT NACH KURSLEVEL UND SILBENMETHODE	143
Abb. 26: GESAMTLERNERFOLG UND NORMALVERTEILUNG	148
Abb. 27: GESAMTLERNERFOLG DIFFERENZIERT NACH INSTITUTION	150
Abb. 28: GESAMTLERNERFOLG DIFFERENZIERT NACH FREMDSPRACHENKENNTNISSEN	152
Abb. 29: GESAMTLERNERFOLG DIFFERENZIERT NACH BERUFSLEHRE	153
Abb. 30: GESAMTLERNERFOLG DIFFERENZIERT NACH DEUTSCHGEBRAUCH	154
Abb. 31: GESAMTLERNERFOLG DIFFERENZIERT NACH VORGÄNGIG BESUCHTEM DEUTSCHKURS	155

Tabellenverzeichnis

TAB. 1: SCHRIFTSPRACHERWERBSMODELL NACH FRITH.....	13
TAB. 2: SCHRIFTSPRACHERWERBSMODELL ERWACHSENER	14
TAB. 3: ZWEITSCHRIFTLERNERKURSE.....	53
TAB. 4: KURSZEITEN INSTITUTION P UND Q	53
TAB. 5: MITTELWERTE DER ZWEITSCHRIFTLERNERKURSE.....	54
TAB. 6: LEHRPERSONEN INSTITUTION P	58
TAB. 7: LEHRPERSONEN INSTITUTION Q	58
TAB. 8: TESTVERFAHREN	67
TAB. 9: ÄUSSERUNGSLÄNGE ELICITED IMITATION, EINSTUFUNGSTEST	71
TAB. 10: ÄUSSERUNGSLÄNGE ELICITED IMITATION, FOLLOW-UP 1	71
TAB. 11: ÄUSSERUNGSLÄNGE ELICITED IMITATION, FOLLOW-UP 2	72
TAB. 12: ÄUSSERUNGEN ELICITED IMITATION, EINSTUFUNGSTEST	74
TAB. 13: ÄUSSERUNGEN ELICITED IMITATION, FOLLOW-UP 1.....	74
TAB. 14: ÄUSSERUNGEN ELICITED IMITATION, FOLLOW-UP 2.....	75
TAB. 15: BEWERTUNGSRASTER ELICITED IMITATION	75
TAB. 16: KUNSTWÖRTER MOTTIER-TEST.....	79
TAB. 17: BEWERTUNGSBOGEN KUNSTWÖRTER.....	80
TAB. 18: ORF-TEXT, EINSTUFUNG.....	84
TAB. 19: ORF-TEXT, FOLLOW-UP 1.....	84
TAB. 20: ORF-TEXT, FOLLOW-UP 2.....	84
TAB. 21: ÄUSSERUNGEN DIKTAT, EINSTUFUNG	95
TAB. 22: ÄUSSERUNGEN DIKTAT, FOLLOW-UP 1.....	96
TAB. 23: ÄUSSERUNGEN DIKTAT, FOLLOW-UP 2.....	97
TAB. 24: BEWERTUNGSRASTER DIKTAT	98
TAB. 25: BEWERTUNGSRASTER LERNERFRAGEBOGEN.....	100
TAB. 26: ITEMANALYSE DIKTAT, ERPROBUNG 1	103
TAB. 27: VERSUCHSPLAN	107
TAB. 28: LEHR- UND LERNUMGEBUNGSVARIABLEN	108
TAB. 29: UNTERRICHTSBEZOGENE VARIABLEN	109
TAB. 30: PERSONENBEZOGENE VARIABLEN	111
TAB. 31: MITTELWERT DER ITEMS, ELICITED IMITATION, EINSTUFUNG	123
TAB. 32: MITTELWERT DER ITEMS, DIKTAT, EINSTUFUNG	124
TAB. 33: ELICITED IMITATION, KORRELATION DER RETEST-ITEMS	125
TAB. 34: DIKTAT, KORRELATION DER RETEST-ITEMS	125
TAB. 35: KUNSTWÖRTER, KORRELATION DER RETEST-ITEMS	125
TAB. 36: HÖRVERSTEHEN, KORRELATION DER RETEST-ITEMS	126
TAB. 37: KORRELATION FORSCHERGENERIERTE TESTS ZU TELC	128

TAB. 38: EFFEKTSTÄRKE NACH COHEN	129
TAB. 39: ÜBERSICHTSTABELLE STATISTISCHE ANALYSEN TELC	130
TAB. 40: TELC-PROZENTWERTE DIFFERENZIERT NACH KURSLEVEL UND SILBENMETHODE	141
TAB. 41: TELC-PROZENTWERTE UNTERTEILT NACH FERTIGKEIT DIFFERENZIERT NACH KURSLEVEL UND SILBENMETHODE	142
TAB. 42: REGRESSIONSMODELL TELC-PROZENTWERTE	144
TAB. 43: ÜBERSICHTSTABELLE STATISTISCHE ANALYSEN GESAMTLERNERFOLG.....	149
TAB. 44: GESAMTLERNERFOLG DIFFERENZIERT NACH KURSLEVEL.....	150
TAB. 45: MITTELWERTE DES LERNERFOLGS ZU ALLEN DREI MESSZEITPUNKTEN.....	155
TAB. 46: LERNERFOLG ERSTE PHASE	156
TAB. 47: LERNERFOLG ZWEITE PHASE	156
TAB. 48: GESAMTLERNERFOLG.....	156
TAB. 49: LERNERFOLG ERSTE PHASE, ELICITED IMITATION.....	157
TAB. 50: LERNERFOLG ZWEITE PHASE, ELICITED IMITATION.....	157
TAB. 51: GESAMTLERNERFOLG, ELICITED IMITATION	157
TAB. 52: LERNERFOLG ERSTE PHASE, HÖRVERSTEHEN	158
TAB. 53: LERNERFOLG ZWEITE PHASE, HÖRVERSTEHEN	158
TAB. 54: GESAMTLERNERFOLG, HÖRVERSTEHEN.....	158
TAB. 55: LERNERFOLG ERSTE PHASE, ORF.....	158
TAB. 56: LERNERFOLG ZWEITE PHASE, ORF.....	159
TAB. 57: GESAMTLERNERFOLG, ORF.....	159
TAB. 58: LERNERFOLG ERSTE PHASE, DIKTAT	159
TAB. 59: LERNERFOLG ZWEITE PHASE, DIKTAT.....	159
TAB. 60: GESAMTLERNERFOLG, DIKTAT	160
TAB. 61: REGRESSIONSMODELL GESAMTLERNERFOLG	161
TAB. 62: GESAMTLERNERFOLG DIFFERENZIERT NACH KURSLEVEL.....	163
TAB. 63: KORRELATION FORSCHERGENERIERTE TESTS DIFFERENZIERT NACH MESSZEITPUNKT UND TESTVERFAHREN	164
TAB. 64: KURSLEVEL UND TREATMENT.....	168
TAB. 65: KURSPHASEN UND MESSZEITPUNKTE	168

Einleitung

Auch wenn es in der Geschichte der Schweiz immer wieder Einwanderungsschübe gab – zu denken wäre beispielsweise an die Glaubensflüchtlinge aus Frankreich, Italien und England Mitte des 16. Jahrhunderts oder, rund hundert Jahre später, an die Fluchtbewegung der Hugenotten nach der Aufhebung des Edikts von Nantes –, so galt die Schweiz lange als klassisches Auswanderungsland. Schweizer migrierten meist auf der Suche nach Arbeit. Der wirtschaftliche Aufschwung im 19. Jahrhundert brachte die Kehrtwende und allmählich begannen mehr Personen aus dem Ausland in die Schweiz einzuwandern als umgekehrt. Betrug die Zahl migrierter Personen 1850 noch 70'000, so war sie bis 1914 auf 600'000 angestiegen. Erst seit Beginn des letzten Jahrhunderts kann die Schweiz somit als typisches Einwanderungsland gesehen werden (vgl. Caroni et al. 2014, S. 23–32).

Im Umgang mit dieser Zuwanderung war die Schweizer Politik bis zu Beginn des ersten Weltkriegs von einer ausgesprochenen Freizügigkeit geprägt. Die aufkeimenden Debatten über die ‚Überfremdung‘ und der Ausbruch des ersten Weltkrieges setzten dem jedoch ein Ende und fortan wurden Einreise und Aufenthalt streng kontrolliert. Diese politische Haltung blieb über Jahrzehnte bestehen. Die Migrationspolitik beinhaltete über die längste Zeit keine Integrationspolitik, sondern beschränkte sich auf Aspekte der Zulassung und Begrenzung. Zum Ende des letzten Jahrhunderts hin änderte sich diese Haltung und Integrationsaspekte traten in den Vordergrund. Auf Ebene der Legislative manifestierte sich dies schlussendlich in der Inklusion eines Integrationsartikels im Bundesgesetz über Aufenthalt und Niederlassung der Ausländer am 1. Oktober 1999 (vgl. ebd.).

Ausländer-, migrations- und nicht zuletzt integrationspolitische Fragen sind spätestens seit 2015, mit dem Beginn einer der grössten Flüchtlingsbewegungen der letzten Jahrzehnte (vgl. UNHCR 2016, S. 2), wieder sehr aktuell. Im Rahmen dieser Migrationsströme haben auch zahlreiche Personen eine Niederlassungs- und Aufenthaltsbewilligung in der Schweiz beantragt, was die Debatte über die Integrationspolitik neu angefacht hat. Es ist nicht Ziel der vorliegenden Dissertation, die Schweizer Integrationspolitik zu kommentieren. Fest steht aber, dass Sprache ein wichtiger Eckpfeiler der Integration sein kann, da erst eine gemeinsame Sprache ein Interagieren der Zugewanderten mit der Aufnahmegesellschaft ermöglicht (vgl. Caroni et al. 2014, S. 16).

In Anbetracht der politischen Relevanz der Sprachförderung lässt sich feststellen, dass gewisse Themen im akademischen Diskurs bisher kaum Fuss gefasst haben. Die Sprachförderung und speziell die Alphabetisierung von Migranten¹ sind zwar von zunehmender Bedeutung in der Zweitspracherwerbsforschung, doch Zweitschriftlernende, wie sie in dieser Dissertation untersucht werden, werden von der Forschung weitgehend vernachlässigt. Dies, obwohl empirische Untersuchungen zeigen, dass die (Vor-)Bildung den Zweitspracherwerb (L2-Erwerb) begünstigt (vgl. Esser 2006a, S. 109) und sich Zweitschriftlernende somit durch ihre Schulgewohntheit und Schriftkenntnisse in der Erst(schrift)sprache (s. Abschn. 1.1.2) deutlich von Analphabeten unterscheiden. Das Potenzial von Zweitschriftlernenden wird im akademischen und politischen Diskurs fast komplett ignoriert, auch in der Schweiz, obwohl anzunehmen ist, dass viele Flüchtlinge Zweitschriftlernende sind. Die Asylstatistik des zweiten Quartals 2019 der Schweiz zeigt, dass Personen aus Eritrea und Syrien mehr als die Hälfte aller anerkannten Flüchtlinge ausmachen (vgl. Staatssekretariat für Migration SEM 2019a, S. 9). Wenn die Literalitätsquoten beider Länder betrachtet werden (81 % im Jahr 2004 in Syrien und 65 % im Jahr 2008 in Eritrea), die über dem Durchschnitt der jeweiligen Region liegen (vgl. The World Bank 2019), kann davon ausgegangen werden, dass die Mehrheit dieser Personen Zweitschriftlernende sind.

Die derzeitige Bildungslandschaft in der deutschsprachigen Schweiz sieht für Zweitschriftlernende vor, dass sie in Alphabetisierungskursen mit Analphabeten unterrichtet werden und dadurch bis zum Übertritt in einen regulären A1-Kurs bis zu drei Jahre in einem solchen Kurs verbringen. Dabei verfügen Zweitschriftlernende im Vergleich zu Analphabeten über mehr Vorwissen und Können in einer Sprache (vgl. Feldmeier 2010a, S. 20), da sie in einem nichtlateinischen Schriftsystem alphabetisiert sind und oft mehrere Jahre Schulerfahrung aufweisen. Deswegen können sie im Normalfall auf zahlreiche Lernstrategien zurückgreifen, erzielen dadurch grössere Fortschritte als Analphabeten im Kursbesuch (vgl. Feldmeier 2015, S. 41) und könnten viel schneller in das reguläre Kurssystem eingegliedert werden, weil sie meist nur wenig Zeit benötigen, um das lateinische Schriftsystem zu erlernen (vgl. Bigelow/Lovrien Schwarz 2010, S. 10). In einem eigens für sie konzipierten Kurskonzept würden sie durch das Anknüpfen an bereits vorhandenen Lese- und

¹ In dieser Dissertation wird aus Gründen der Lesbarkeit auf eine durchgehende Benutzung geschlechtsneutraler Bezeichnungen oder eine konsequente Verwendung der weiblichen und männlichen Form verzichtet. Alle männlichen oder weiblichen Rollenbezeichnungen gelten im Sinne der Gleichbehandlung grundsätzlich für beide Geschlechter.

Schreibkompetenzen eine bedarfsgerechte und zielgerichtete Förderung erhalten, wie dies im *Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* (BAMF 2018a) empfohlen wird. Dadurch wäre auch eine schnellere Eingliederung in den Arbeitsmarkt oder in die Ausbildung möglich (vgl. BAMF 2017a, S. 3–4). Es ist davon auszugehen, dass im aktuellen Bildungssystem der Schweiz Zweitschriftlernende einerseits viel Zeit bis zum Übertritt in einen regulären A1-Niveaunkurs verlieren und andererseits ihre Sprachpauschale (sie umfasst oftmals einen fixen Betrag für Deutschkurse, s. Abschn. 2.2.2) ineffizient eingesetzt wird. Häufig ist diese bis zum Übertritt in einen A1-Kurs vollständig aufgebraucht, sodass dieser Deutschkurs nicht besucht werden kann oder mit anderen Mitteln finanziert werden muss. Anzunehmen ist auch, dass die Motivation der Teilnehmenden nachlässt, da das Tempo für Zweitschriftlernende in regulären Alphabetisierungskursen zu gering ist (vgl. Schuller et al. 2012, S. 32).

In Deutschland wurde dieses Defizit am Kurssystem erkannt und im Februar 2017 mit den ersten *Zweitschriftlernerkursen* begonnen, die anhand der Vorgaben des *Konzepts für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* unterrichtet werden. Ziel dieser Kurse ist eine vollständige Einführung in das lateinische Schriftsystem innerhalb der ersten 100 Kurslektionen und das Erreichen des Niveaus A1 nach dem Gemeinsamen europäischen Referenzrahmen (GER) innerhalb weiterer 200 Kurslektionen. Die Kurse werden anschliessend fortgeführt und Teilnehmende sollen nach zusätzlich maximal 600 Unterrichtseinheiten das Niveau B1 nach GER erreichen, insgesamt innerhalb von 900 Kurslektionen.

Auch in Österreich ist diesbezüglich ein Manko wahrgenommen worden, sodass im Jahr 2016 ein Pilotprojekt mit Zweitschriftlernenden durchgeführt wurde. Die Schweiz hingegen hinkt im Bereich der Alphabetisierung von Migrant*innen hinterher – sie besitzt weder ein gesondertes Konzept für Alphabetisierungs- oder Zweitschriftlernerkurse noch verbindliche Richtlinien dafür. Institutionen wie auch Lehrpersonen können dementsprechend die Alphabetisierungskurse so handhaben, wie sie es für richtig erachten.

Auch wenn sich die Thematik Zweitschriftlernende bisher in der Forschung noch nicht etabliert hat, verweist bereits das Rahmencurriculum von Lenz et al. (2009) darauf, dass Zweitschriftlernende nicht zusammen mit Analphabeten einen Deutschkurs besuchen sollten, vor allem dann nicht, wenn dadurch schulgewohnte Lernende mit Schulungsgewohnten zusammen unterrichtet werden. Laut den Autoren benötigen Zweitschriftlernende, im

Gegensatz zu Analphabeten, lediglich einen Vorkurs, um mit den Materialien und Prinzipien im eigentlichen Sprachkurs vertraut zu werden (vgl. Lenz et al. 2009, S. 15, 33). In der Schweiz wird demzufolge zumindest von manchen Sprachwissenschaftlern seit einiger Zeit eine Trennung von Analphabeten und Zweitschriftlernenden vorgeschlagen, auch wenn sich diese in der Praxis nicht etabliert hat.

Die Silbenmethode wird in der Literatur zum Teil zwar kontrovers diskutiert, gilt jedoch als wichtige Methode für den Lese- und Schreiberwerb des Deutschen. Dies zeigt sich auch darin, dass die Silbenmethode in Lehrwerken und in der Alphabetisierungsarbeit seit einiger Zeit fast omnipräsent ist. Mit Zweitschriftlernenden empirisch untersucht wurde sie aber bisher nur von Albert et al. (2015), die der Silbenmethode einen positiven Nutzen attestieren. Der Bereich Hörverstehen und Sprechen bleibt jedoch auch bei ihnen unerforscht, weshalb es bis heute an einer umfassenden Studie zur Silbenmethode mit Zweitschriftlernenden oder Analphabeten fehlt.

Die vorliegende Qualifizierungsarbeit möchte dazu einen Forschungsbeitrag leisten und einen Denk- und Diskussionsanstoss liefern, indem sie den Einsatz von Zweitschriftlernerkursen mit oder ohne speziellem Fokus auf der Silbenmethode untersucht. Daraus ergeben sich folgende Forschungsfragen:

- F1. Erreichen Zweitschriftlernende, die in der deutschsprachigen Schweiz nach dem Deutschen Zweitschriftlernerkonzept unterrichtet werden, innerhalb von 300 Kurslektionen ein A1-Niveau nach telc?
- F2. Welche Unterschiede sind in verschiedenen Teilbereichen der Sprachkompetenz feststellbar, wenn Zweitschriftlernende im Rahmen des Zweitschriftlernerkonzepts aus Deutschland mit oder ohne speziellem Fokus auf der Silbenmethode unterrichtet werden?

Ziel dieser Forschungsarbeit ist einerseits die Überprüfung der Wirksamkeit des *Konzepts für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* und andererseits die Überprüfung der Wirksamkeit der Silbenmethode. Zu diesem Zweck wurde mit einer Interventionsstudie (Feldexperiment) gearbeitet. Damit der Einfluss des *Zweitschriftlernerkonzepts* und der Silbenmethode auf die sprachliche Kompetenz sichtbar gemacht werden kann, wurden Erhebungen zu verschiedenen Messzeitpunkten durchgeführt: ein Einstufungstest vor Kursstart, ein Follow-up-Test nach 100 Kurslektionen und einer nach 200 Kurslektionen sowie die standardisierte A1-Prüfung von ‚The European Language Certificates‘ (telc) nach 300 Kurslektionen. Da nicht auf ein bereits vorhandenes Testverfahren zurückgegriffen werden konnte, wurde eigens für die Forschungsarbeit ein ca. 50-minütiges

Testverfahren entwickelt, das sich an der allgemeinen Sprachkompetenz nach dem Modell von Bachman und Palmer (1990) orientiert. Das Testverfahren umfasst insgesamt fünf Testinstrumente: Elicited Imitation für das Sprechen, Kunstwörter für das Arbeitsgedächtnis, Oral Reading Fluency für die Leseflüssigkeit, Diktat für die Laut-Buchstaben-Zuordnung und die Verständlichkeit sowie ein Hörverstehenstest. Diese Testinstrumente dienen dazu, sprachliche Leistungsunterschiede zwischen den Treatmentgruppen (*Zweitschriftlernerkonzept* mit/ohne Einsatz der Silbenmethode) zu analysieren.

Die Arbeit ist in insgesamt vier Teile gegliedert: theoretische Grundlagen, Methodologie, Ergebnisse und Diskussion.

Teil 1: Grundlagen

Im ersten Kapitel werden die konzeptuellen und empirischen Grundlagen erläutert und einige wichtige Schlüsselbegriffe definiert. So soll beispielsweise klar zwischen der Erstsprache und der Erstschriftsprache unterschieden werden, denn diese können, müssen aber nicht deckungsgleich sein. Der erste Abschnitt widmet sich anschliessend der Definition der Zielgruppe der Zweitschriftlernenden, bevor erklärt wird, woher die Silbenmethode ursprünglich stammt und wie sie im Zweitsprachenunterricht mit erwachsenen Migranten angewendet wird.

Im zweiten Abschnitt wird auf das deutsche Kurskonzept für Zweitschriftlernende eingegangen: von der Einstufung der Teilnehmenden, über den Kursverlauf bis zum Abschlusstest. Anschliessend sollen kontrastierend die Gegebenheiten in der Schweiz, spezifischer im deutschsprachigen Teil des Kantons Bern, gegenübergestellt werden.

Teil 2: Methodologie

Im zweiten Teil der Arbeit wird als Erstes der Forschungsgegenstand vorgestellt (s. Kap. 3). Dazu gehören die Unterscheidung der Kontroll- und Experimentalgruppe sowie Details zur Stichprobenbildung. Bei empirischen Untersuchungen mit einer als vulnerabel geltenden Zielgruppe ergeben sich zusätzliche Herausforderungen, auf die näher eingegangen wird. Am Ende des dritten Kapitels werden dann die zwei Forschungsfragen sowie die sieben Hypothesen samt Unterhypothesen vorgestellt, die im Rahmen der vorliegenden Studie untersucht werden.

Das vierte Kapitel wird mit einigen Überlegungen zur Unterscheidung zwischen kommunikativer und allgemeiner Sprachkompetenz eingeleitet, bevor auf die in der Studie

verwendeten forschergenerierten Tests einzeln eingegangen wird. Bei der Beschreibung der Tests werden das Testkonstrukt, die jeweiligen Testaufgaben und die Testbewertung erklärt. Anschliessend wird auf die Datengrundlage eingegangen (Kap. 5). Wie die Tests erprobt und überarbeitet wurden und wie der genaue Prüfungsablauf aussah, wird am Anfang dieses Kapitels dargestellt. Anschliessend wird auf die Kodierung der Variablen eingegangen. Dabei werden die abhängigen und die unabhängigen Variablen genauer erklärt, woraufhin einige Bemerkungen zur Datenaufbereitung folgen.

Teil 3: Ergebnisse

Kapitel 6 ist das Kernstück der Studie und stellt die Ergebnisse vor, die im Rahmen der Dissertationsarbeit gewonnen wurden. Bevor auf statistische Ergebnisse eingegangen werden kann, wird die Testgüte der selbst erstellten Tests vorgestellt. Mittels einer klassischen Itemanalyse wird auf die drei Gütekriterien Objektivität, Reliabilität und Validität eingegangen (s. Abschn. 6.1).

Zur Überprüfung der beiden Forschungsfragen und der Hypothesen wurden verschiedene statistische Analysen durchgeführt. In Abschnitt 6.2 wird zuerst auf die Resultate aus der telc-Prüfung eingegangen, bevor die korrespondierenden Hypothesen und Unterhypothesen besprochen werden. Analog dazu werden in Abschnitt 6.3 zuerst die Daten aus den selbst generierten Tests analysiert, bevor auch hier die entsprechenden Hypothesen samt Unterhypothesen betrachtet werden.

Der letzte Abschnitt widmet sich der statistischen Überprüfung der Kunstwörter und dient der Überprüfung, ob der Unterricht mit der Silbenmethode die Kapazität des Arbeitsgedächtnisses steigert und inwiefern sich das Arbeitsgedächtnis positiv auf die vier Fertigkeiten Sprechen, Hören, Schreiben und Lesen auswirkt (s. Abschn. 6.4).

Teil 4: Diskussion

In Kapitel 7 werden die Ergebnisse anhand der zwei Forschungsfragen sowie anhand der sieben Hypothesen diskutiert. An dieser Stelle wird auch Platz für weitere Ergebnisse geschaffen, die nicht durch die Forschungsfragen oder Hypothesen abgedeckt werden.

Anschliessend wird auf einige Limitationen der vorliegenden Studie näher eingegangen, welche die Testinstrumente, das Forschungsdesign und die externen Forschungspartner betreffen (Kap. 8). Auch soll kritisch reflektiert werden, inwiefern die Befunde ausserhalb des Testsettings, auf die restliche Schweiz und auf unsere Nachbarländer generalisierbar sind.

In Kapitel 9 werden die Implikationen der Dissertation für die weitere Forschung, die Bildungspolitik und die Praxis diskutiert. Ausser Frage steht, dass der Status quo nicht zielführend ist und gezielt Kurse für Zweitschriftlernende entwickelt werden müssen.

Abgerundet wird die Arbeit mit einem Schlusswort (Kap. 10).

Anhänge

Die Dissertation beinhaltet drei Anhänge: den Anhang am Ende der vorliegenden Arbeit, einen separat gedruckten und in der Bibliothek der Universität Fribourg ausleihbaren Anhang und einen zusätzlichen Anhang auf CD-ROM. Der Anhang am Ende der Arbeit umfasst die Dokumente, die für das unmittelbare Verständnis der Dissertation relevant sind. Dazu gehören ein Beispiel für ein Silbenprotokoll, Unterrichtsbeispiele mit der Silbenmethode, die Diktatvorlage, die Vorlagen für die Hörverstehenstests sowie der Lehrer- und Lernerfragebogen (deutsche Version). Im separat gebundenen Anhang sind weitere Dokumente, u. a. die übersetzten Versionen des Lernerfragebogens, die Transkriptionen des Elicited Imitation und die Prüfungshandreicherung, sowie alle statistischen Auswertungen vorzufinden. Der Aufbau und Inhalt des separaten Anhangs entspricht der Struktur der vorliegenden Arbeit. Der dritte Anhang auf CD-ROM umfasst schlussendlich die kompletten SPSS-Dateien, die Verschriftlichung der Antworten der Teilnehmenden bei den Diktaten sowie die Audiodateien, die bei den Tests zum Einsatz kamen.

Teil I: Grundlagen

1. Konzeptuelle und empirische Grundlagen

In diesem Teil der Arbeit werden die für das behandelte Thema relevanten theoretischen Grundlagen vorgestellt. Dazu werden zuerst Schlüsselbegriffe definiert (s. Abschn. 1.1), bevor der Zweitschifterwerbsprozess bei Erwachsenen behandelt wird (s. Abschn. 1.2). Anschliessend wird die Zielgruppe beschrieben (s. Abschn. 1.3) und näher auf die Silbenmethode eingegangen (s. Abschn. 1.4).

1.1 Schlüsselbegriffe

Seit längerer Zeit sind nicht nur die Begriffe der Alphabetisierung bzw. des Analphabetismus, sondern auch die Literalität und die Grundbildung Themen der fachlichen Diskussion (s. z. B. Rokitzki 2016, S. 29). Je nach Sprache divergieren die verwendeten Definitionen stark, weshalb eine genauere Betrachtung und Eingrenzung dieser Begriffe sinnvoll ist. Auf einen historischen Überblick zum Bedeutungswandel der Begriffe wird jedoch verzichtet.²

1.1.1 Flüchtlinge

In dieser Qualifizierungsarbeit werden Zweitschriftlernende mit Fluchthintergrund untersucht. Seit der Genfer Flüchtlingskonvention von 1952 wird eine Person als Flüchtling³ definiert, wenn sie sich

[...] aus begründeter Furcht vor Verfolgung wegen ihrer Rasse, Religion, Staatszugehörigkeit, Zugehörigkeit zu einer bestimmten sozialen Gruppe oder wegen ihrer politischen Überzeugung ausserhalb ihres Heimatlandes befindet und dessen Schutz nicht beanspruchen kann oder wegen dieser Befürchtungen nicht beanspruchen will [...] (Der Bundesrat 2018).

Anders als Migranten, die häufig versuchen, ihre wirtschaftliche Situation zu verbessern und deswegen aus freiem Willen auswandern, sind Flüchtlinge aus verschiedenen Gründen gezwungen, ihr Heimatland zu verlassen (vgl. Amnesty International Schweiz o. J.).

Bei empirischen Untersuchungen mit einer als vulnerabel geltenden Gruppe können sich Herausforderungen ergeben, die über die bekannten methodischen und/oder praktischen Schwierigkeiten empirischer Forschungen hinausgehen (vgl. Röder et al. 2018, S. 314). In

² Für eine detaillierte Übersicht siehe Literaturbericht Alphabetisierung erwachsener Migrant/innen (Guerrero Calle et al. im Erscheinen).

³ Für diese Qualifizierungsarbeit wurde der Begriff *Flüchtling* gewählt, wie er sich in der juristischen Fachsprache etabliert hat. Hier soll jedoch explizit angemerkt werden, dass der Begriff frei von politischen Sichtweisen oder mit dem Begriff verbundenen negativen Haltungen zu verstehen ist.

Abschnitt 3.4 wird näher auf die Zielgruppe und die damit verbundenen potenziellen, methodischen und praktischen Herausforderungen eingegangen.

1.1.2 Erstsprache und Erstschriftsprache

Aus monolingualer Tradition wird im sprachwissenschaftlichen Gebrauch mit der Erstsprache, wie der Name suggeriert, die erste erlernte Sprache bezeichnet: Es ist die Sprache, die ein Kind konkret erlernt und in der es gleichzeitig auch das Sprechen und Sprache an sich lernt. Die Erstsprache ist prägend für die Weiterentwicklung eines Kindes, da sie Grundlage mentaler Konzepte und Denkprozesse ist. Die Erstsprache prägt die Sozialisierung und ist oft durch eine emotionale Bindung gekennzeichnet (vgl. Olariu 2007, S. 304–305).

Der Begriff Muttersprache wird häufig synonym zur Erstsprache verwendet, wobei die emotionale Komponente, die Beziehung zur Mutter, stärker in den Vordergrund tritt. Die Verwendung des Begriffs in linguistischen Studien wird aber kritisiert, da er teils unreflektierte umgangssprachliche, normative oder gar nationalistische Konnotationen trägt (vgl. König 2016, S. 271–274). Zudem ist der Begriff «für Sprachlehr- und -lernzusammenhänge nicht gut geeignet, weil die Erstsprache keineswegs immer die Sprache der Mutter ist» (Hufeisen/Riemer 2010, S. 738).

Da die vorliegende Arbeit *Zweitschriftlernende* untersucht, muss zudem der Begriff der *Erstschriftsprache* eingeführt werden. Häufig entspricht die Erstsprache der Sprache, in der ein Individuum als Erstes mündliche, aber auch schriftliche Fertigkeiten aufbaut. Bei einigen Zweitschriftlernenden ist dies aber nicht der Fall. Bei Kurdisch sprechenden Personen ist beispielsweise häufig zu beobachten, dass diese zwar Kurdisch als ihre Erstsprache bezeichnen, jedoch nur auf Arabisch schreiben und lesen gelernt haben. Ähnliches kann auch bei Personen aus Eritrea festgestellt werden, die beispielsweise fließend Tigrinya sprechen und eine emotionale Bindung zu dieser Sprache aufweisen, jedoch in Koranschulen das Lesen und Schreiben auf Arabisch gelernt haben, weshalb auch hier die Erstsprache und die Erstschriftsprache nicht deckungsgleich sind.

In dieser Arbeit wird somit – wie in der Linguistik üblich (vgl. Goethe Institut 2009) – der Begriff Erstsprache verwendet, wenn von derjenigen Sprache die Rede ist, in der ein Teilnehmer sozialisiert wurde und die er mehrheitlich in seinem familiären und persönlichen sozialen Umfeld spricht. Synonym hierzu wird auch L1 verwendet.

Wenn jemand in der Erstsprache keine schriftlichen Kenntnisse aufweist, jedoch in einer anderen Sprache lesen und schreiben kann, dann wird zudem von der Erstschriftsprache die Rede sein.

1.1.3 Analphabetismus, Grundbildung und Literalität

In den siebziger Jahren wurde in den Industrieländern die Problematik des Analphabetismus erstmals thematisiert. Die daraus entstandene Diskussion um präzise Definitionen hält bis heute an (vgl. Löffler/Korfkamp 2016, S. 9; Rokitzki 2016, S. 29). Im deutschsprachigen Raum prägen Begriffe wie Alphabetisierung und Grundbildung die öffentlich-mediale Diskussion (vgl. Tröster/Schrader 2016, S. 42).

Der Begriff **Analphabetismus** wird als mangelhaftes oder gänzlich fehlendes Beherrschen der Schriftsprache definiert. Dabei können verschiedene Erscheinungsformen unterschieden werden, die auf der zeitlichen Ebene und der Kenntnisebene differieren. Auf der Zeitebene unterscheidet Linde (2007, S. 51, 91–92) den primären Analphabetismus und den sekundären Analphabetismus. Betroffene des Ersteren verfügten nie über Schriftsprachkenntnisse. Sekundäre Analphabeten haben hingegen in der obligatorischen Schulzeit lesen und schreiben gelernt, diese Fähigkeiten jedoch im Laufe des Lebens verloren, z. B. aufgrund mangelnder Anwendung. Auf der Kenntnisebene unterscheidet die Autorin den totalen Analphabetismus und den funktionalen Analphabetismus. Demnach verfügen totale Analphabeten über keinerlei Schriftsprachkenntnisse, während funktionale Analphabeten (in der Schweiz ist dafür der Begriff ‚Illettrismus‘ gebräuchlich) zwar Schriftsprachkenntnisse haben, diese jedoch den gesellschaftlichen Mindestanforderungen nicht genügen (vgl. Linde 2008, S. 92–93; s. auch Döbert/Hubertus 2000, S. 20, 23; Tröster/Schrader 2016, S. 43, 45).⁴ Werden die beiden Ebenen übereinandergelegt, wird ersichtlich, dass primäre Analphabeten zugleich als totale Analphabeten klassifiziert werden können.

Burt und Peyton, deren Definitionsansatz besonders im englischsprachigen Raum verbreitet ist, klassifizieren Analphabetismus nicht primär aufgrund der Kenntnisse oder auf der zeitlichen Ebene wie Linde, sondern anhand des Zugangs einer Person zur Schriftlichkeit. So unterscheiden sie unter anderem zwischen ‚Nonliterate‘, die aus schriftlichen Kulturen stammen, aber wegen des fehlenden Bildungszugangs weder im Kindes- noch im Jugendalter

⁴ Für eine detaillierte Erklärung der verschiedenen Erscheinungsformen siehe Burt und Peyton (2003).

lesen und schreiben gelernt haben, und ‚Preliterates‘, die Personen aus schriftlosen Kulturen umfassen (vgl. Burt/Peyton 2003, S. 5).

Bisher existiert auch keine allgemeingültige Definition für **Grundbildung**. Allen Definitionsversuchen ist jedoch gemeinsam, dass eine Vielzahl von Fertigkeiten genannt werden, die für eine aktive Partizipation an der Gesellschaft nötig sind (vgl. z. B. Linde 2008, S. 53–54; Tröster/Schrader 2016, S. 45). Grundbildungsangebote sollen eine gezielte Verbesserung der Lernfähigkeit und die Vermittlung von Lern- und Problemlösungsstrategien beinhalten (vgl. z. B. Derichs-Kunstmann et al. 1993, S. 61; Linde 2008, S. 50; Hammann et al. 2013, S. 21). Damit umfasst Grundbildung demzufolge mehr als das Lernen der Schriftsprache, die nicht genügt, um im heutigen Alltag zurechtzukommen (vgl. Tröster/Schrader 2016, S. 46). Während unsere deutschen Nachbarstaaten die Grundbildung bereits seit mehreren Jahren in ihren Curricula für fremdsprachige Erwachsene verortet haben (vgl. Fritz et al. 2006, S. 27; Feldmeier 2007, S. 8, 2015, S. 18), ist dies in der Schweiz bisher nicht der Fall. Zugespitzt kann damit gesagt werden, dass in regulären Alphabetisierungskursen in der deutschsprachigen Schweiz lesen und schreiben gelernt wird, jedoch ohne Rücksicht auf Lern- und Problemlösungsstrategien.

Ähnlich wie die Grundbildung schliesst **Literalität** nebst schriftsprachlichen Kompetenzen auch grundlegende Rechenfähigkeiten und Allgemeinwissen mit ein und bezieht sich auf eine elementare Schriftkultur (vgl. Rokitzki 2016, S. 27–28). Anders ausgedrückt umfasst Literalität den sozialen Aspekt des Lesen- und Schreibenkönnens – Alphabetisiertheit wäre dazu der technische Aspekt, der jedoch allein nicht ausreicht, um an einer Schriftkultur teilzuhaben (vgl. Panagiotopoulou 2001, S. 55). Die Definitionen von Grundbildung und Literalität beziehen sich somit immer auf einen sozialen Kontext, in dem sich die Personen bewegen. Literalität streicht jedoch im Gegensatz zur Grundbildung die Individualität dieser Personen heraus: Das Konzept umfasst nicht nur die «Verwendung von gedruckten und geschriebenen Informationen, um in der Gesellschaft zurechtzukommen», sondern auch die Möglichkeit, «eigene Ziele zu erreichen und eigenes Wissen sowie die individuellen Möglichkeiten zu entwickeln» (OECD 1998, S. 16).

Alphabetisierungskurse sollten laut konzeptionellen Vorstellungen somit mehr umfassen als rein technische Aspekte des Lesen- und Schreibenkönnens. Sie sollten die Grundbildung und Aspekte der Literalität einbeziehen, damit Lernende an der Gesellschaft teilhaben, Wissen entwickeln und eigene Ziele verfolgen können. Ein solches Kurssystem umfasst demzufolge

das Lesen- und Schreibenlernen, die Rechenfähigkeit, den Aufbau des Allgemeinwissens, die Verbesserung der Lernfähigkeit und die Vermittlung von Lern- und Problemlösungsstrategien. Obwohl die Theorie z. T. bereits mehr als zehn Jahre alt ist, werden diese Gesichtspunkte in vielen deutschsprachigen Alphabetisierungskursen in der Schweiz nicht oder kaum berücksichtigt. Hier zeichnet sich unter anderem ein grosser Forschungsbedarf ab.

1.2 Zweitschrifterwerb

Wie bereits erläutert, werden in Alphabetisierungskursen in der deutschsprachigen Schweiz mehrheitlich die Fertigkeiten des Lesens und Schreibens fokussiert. Diese beiden Kompetenzen sind massgebend für den Übertritt in A1-Kurse (vgl. Hammann et al. 2013, S. 25). Zudem sind die Lese- und Schreibkenntnisse grundlegender Bestandteil der Klassifikation als Analphabet resp. Zweitschriftlerner, weshalb im folgenden Abschnitt lediglich auf den Schriftspracherwerb eingegangen wird, der hier als Lese- und Schreibenlernen verstanden wird.

1.2.1 Schriftspracherwerb

Obwohl in der Zweitschrifterwerbsforschung davon ausgegangen wird, dass Lernende im Schreiben einen weniger grossen Fortschritt als z. B. im Lesen erzielen (vgl. Brooks et al. 2001, S. 77–79; Tranza/Sunderland 2009, S. 22), gibt es bis dato praktisch keine Forschung in Bezug auf den Erst- oder den Zweitschrifterwerb bei Flüchtlingen und Migranten (vgl. Feldmeier 2010a, S. 31; Rackwitz 2016, S. 51). Daher bestehen derzeit lediglich Vermutungen über den Schriftspracherwerb dieser Zielgruppen, auf die im Folgenden näher eingegangen wird.

Der Schriftspracherwerb bei erwachsenen Analphabeten verläuft laut der Schriftspracherwerbstheorie ähnlich wie bei Kindern. Diese Annahme stammt vor allem von Praktikern, die ihre Beobachtungen verschriftlicht haben; es fehlen bisher entsprechende empirische Untersuchungen (vgl. ebd., S. 51–54).

Beim Schriftspracherwerb von Kindern geht die Forschung von drei wesentlichen Stufen bzw. Phasen aus, die an das Modell von Frith (1985) für den kindlichen Erwerbsprozess des Englischen angelehnt sind (vgl. Feldmeier 2010a, S. 29). Obwohl im Laufe der Zeit zahlreiche Entwicklungsmodelle erarbeitet wurden,⁵ wird hier lediglich auf die ursprünglichen Ansätze von Frith und Günther eingegangen, da die moderneren Modelle auf diesen aufbauen.

⁵ Nach Erscheinungsjahren (vgl. Siekmann/Thomé 2018, S. 97–146): Frith (1985), Dehn (1985), Günther (1986), Spitta (1988), Valtin (1988), Scherer-Neumann (1989), May (1990 & 2002), Brügelmann/Brinkmann (1994) und Thomé (2006).

Die erste Stufe des Modells von Frith (1985) ist die sogenannte **logografische Phase**, in der Kinder mit dem Lesen beginnen. Erst nachdem die logografische Strategie im Lesen erworben wurde, beginnen die Kinder auf diese Art und Weise zu schreiben. Häufig werden Wörter zeichnerisch dargestellt, z. B. ☾ für Mond. In dieser ersten Phase geben Kinder vielfach an, das Wort *Kuh* sei länger als das Wort *Regenwurm* – weil eine physische *Kuh* grösser ist als ein *Regenwurm*. Zudem kommt es in dieser Phase vermehrt zu Buchstabenvertauschungen oder Buchstabenauslassungen (vgl. Dürscheid 2016, S. 246).

In der zweiten Stufe des Modells, der **alphabetischen Phase**, liegt die Aufmerksamkeit auf Lauten und Buchstaben (vgl. ebd.). Kinder lesen in dieser Phase Wörter phonographisch, Laut für Laut. Den Zugang zu dieser Phase erhalten Kinder nach Frith (1985) über das Schreiben und nicht über das Lesen, wie dies in der ersten Stufe der Fall ist. Es ist umstritten, ob die phonologische Bewusstheit (s. Abschn. 1.4.2) als Vorläufertätigkeit angesehen werden muss oder ob sie eine Folge des Schriftspracherwerbs ist (vgl. Stetter 2005, S. 108–114).

In der dritten Stufe, der **orthografischen Phase**, lesen Kinder die Wörter ganzheitlich und nicht mehr durch phonologisches Rekodieren. Sie orientieren sich nun an Buchstabenfolgen und Buchstabenkombinationen. In einem zweiten Schritt wird diese Phase auf das Schreiben übertragen: Orthografische Regularitäten werden mehr und mehr umgesetzt, morphologische Schreibungen und die Klein- und Grossschreibung werden beachtet. Die Silbenstruktur wird nicht mehr nur benutzt, um Lauteinheiten richtig zu artikulieren, wie in der logografischen Phase, sondern auch, um Schreibregularitäten zu realisieren (vgl. Dürscheid 2016, S. 248). Die dritte Phase stellt den Abschluss des Schriftspracherwerbs nach Frith (1985) dar.

Zusammengefasst lässt sich das Frithsche Modell folgendermassen visualisieren:

Phase/Stufe	Erster Schritt	Zweiter Schritt
1	Lesen	Schreiben
2	Schreiben	Lesen
3	Lesen	Schreiben

Tab. 1: Schriftspracherwerbsmodell nach Frith (i. A. an Frith 1985)

Wie aus Tab. 1 ersichtlich ist, sind Lesen und Schreiben von wechselnder Wichtigkeit. Es muss aber angemerkt werden, dass Frith (1985) ihr Modell nur aufgrund von Beobachtungen von Leseprozessen erstellt hat und das Schreiben dann anhand entsprechender Überlegungen und nicht aufgrund tatsächlicher Beobachtungen ergänzt hat (vgl. Siekmann/Thomé 2018, S. 118–119).

Ein weiteres Schriftspracherwerbsmodell hat Günther (1986) entwickelt, indem er das Frithsche Modell (1985) um zwei Stufen erweitert und es für die deutsche Sprache adaptiert hat. Dieses Modell beginnt mit einer zusätzlichen Stufe, der präliteral-symbolischen Phase, die als Phase 0 vorangeht. Hier ahmen Kinder Erwachsene beim Lesen und Schreiben nach und benutzen symbolische Darstellungen, z. B. ♥ für Liebe. Diese Stufe wird als Vorbereitung für den Schriftspracherwerb angesehen (vgl. Günther 1986, S. 34–35; Dürscheid 2016, S. 249). Die vierte Phase, die integrativ-automatisierte Phase, die den Abschluss des Erwerbsprozesses bildet, ist eine weitere Modifikation des Frithschen Modells. In dieser Stufe wird eine Sicherheit im Lesen und Schreiben erworben, wie sie für automatisierte Handlungen charakteristisch ist. Diese Stufe stellt eigentlich keine neue Phase dar, sondern entspricht dem schriftlichen Sprachgebrauch von kompetenten Lesern und Schreibern (vgl. Günther 1986, S. 43).

Zu nennen ist darüber hinaus das Schriftspracherwerbsmodell für erwachsene Migranten des österreichischen Autorenteamts Fritz et al. (2006), das einige Parallelen zu dem von Frith (1985) zeigt. Die beiden Prozesse Lesen und Schreiben sind auch hier eng miteinander verbunden, die Phasen werden aber nur durch das Schreiben eingeleitet (s. Tab. 2). Dem Modell fehlt jedoch die systematische wissenschaftliche Validierung, denn es wurde auf der Grundlage von Praxisbeobachtungen in einem vierstufigen Kurssystem entwickelt (vgl. Fritz et al. 2006, S. 4). Inwiefern dieses Modell für die Zielgruppe übernommen werden kann, kann hier mangels einer empirischen Überprüfung nicht abschliessend beantwortet werden.

Phase/Stufe	Erster Schritt	Zweiter Schritt
1	Schreiben	Lesen
2	Schreiben	Lesen
3	Schreiben	Lesen
4	Schreiben	Lesen

Tab. 2: Schriftspracherwerbsmodell Erwachsener (vgl. Fritz et al. 2006, S. 33–38)

Obwohl im Bereich der Schriftspracherwerbsmodelle für Erwachsene bisher keine empirischen Untersuchungen vorliegen, geht die Forschungstheorie davon aus, dass die logografische Phase bei erwachsenen Analphabeten weniger relevant ist als bei Kindern (vgl. Feldmeier 2010b, S. 323; Rackwitz 2016, S. 51). Die alphabetische Phase hingegen wird auch bei Erwachsenen als wesentlich angesehen (vgl. Nickel 1998, S. 21; Hubertus/Nickel 2003, S. 724; Feldmeier 2010b, S. 34; Rackwitz 2016, S. 52). Es muss jedoch angemerkt werden, dass sich die Modelle von Frith und Günther auf den erssprachlichen Schriftspracherwerb beziehen. Inwiefern diese auch für den Zweitspracherwerb Gültigkeit haben, bleibt bis dahin

ungeklärt. Zusammenfassend lässt sich festhalten, dass gerade im Schriftspracherwerb ein beträchtliches Forschungsdesiderat bezüglich erwachsener Analphabeten und erwachsener Zweitschriftlerner besteht.

1.2.2 Transfer

Bevor der Forschungsstand zum Schriftspracherwerb diskutiert werden kann, muss zunächst die Frage nach möglichen Transfers geklärt werden. In Alphabetisierungstheorien wird vielfach angeführt, dass die Erstsprache Einfluss auf den Zweitspracherwerb haben kann (s. z. B. Rokitzki et al. 2013, S. 102; Feldmeier 2015, S. 112; Feick/Schramm 2016, S. 216–217) – es wird demnach von der Kontrastivhypothese ausgegangen. In diesen theoretisch ausgerichteten Arbeiten wird dieser Punkt besonders hinsichtlich der Sprachbewusstheit, des Sprachvergleichs und der kontrastiven Alphabetisierung⁶ besprochen.

Im Fachdiskurs wird von Transfer gesprochen, doch eine übereinstimmende Definition findet sich in der Literatur nicht. Es wird jedoch mehrfach auf Odlin verwiesen, der den Begriff folgendermassen definiert: «Transfer is the influence resulting from similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired» (Odlin 1989, S. 27). Beim Transfer geht es demnach um Ähnlichkeiten und Unterschiede zwischen der Ausgangs- und der Zielsprache. Üblicherweise wird unterschieden zwischen positiven Transfers (Ähnlichkeiten mit Gültigkeit in beiden Sprachen) und negativen Transfers bzw. Interferenzen (Übertragungen aus der Ausgangssprache, die so aber in der Zielsprache nicht existieren und demnach zu Fehlern bzw. Unsicherheiten führen) (vgl. ebd., S. 26).

Die Kontrastivhypothese ist umstritten, da Fehler weder generell durch die Ausgangssprache vorhergesagt werden können noch durch die Erstsprache erklärbar sein müssen. Dennoch ist der didaktische Wert von expliziten Hinweisen auf mögliche Transfers nicht zu unterschätzen, denn bereits gelernte Sprachen können den Zweitspracherwerb beeinflussen und Transfers können als natürliche Erwerbsstrategien angesehen werden (vgl. Kuhberg 2001, S. 655–661). Zusätzlich stellt ein solcher Rückgriff auf die Muttersprache⁷ der Teilnehmenden eine Würdigung dieser Sprache dar (vgl. Feick/Schramm 2016, S. 217).

⁶ In der kontrastiven Alphabetisierung wird die L1 der Teilnehmenden aktiv einbezogen (vgl. Feldmeier 2005a).

⁷ Wie in Kapitel 1.1.2 beschrieben, wird in der vorliegenden Arbeit anstelle der Muttersprache der Begriff Erstsprache (resp. Erstschriftsprache) verwendet. Die hier diskutierte Methode wird dennoch *Rückgriff auf die Muttersprache* genannt, weil sich diese Benennung in der Literatur so etabliert hat.

1.2.3 Forschungsstand Zweitschrifterwerb

Wie im vorherigen Abschnitt in Bezug auf Transfers beschrieben, geht die Theorie zur Alphabetisierung erwachsener Migranten mehrfach davon aus, dass die Ausgangssprache der Teilnehmenden einen positiven Einfluss auf den Zweitspracherwerb haben kann und demnach ein positiver Transfer stattfindet. So konnte beispielsweise Heyn (2013) in ihrer Forschungsarbeit, die an das Projekt *Alphamar* angegliedert ist und in der primäre Analphabeten sowie Zweitschriftlernende untersucht wurden, zeigen, dass ein Rückgriff auf die Muttersprache von Vorteil sein kann, insbesondere, um grammatische Kenntnisse in der Zielsprache zu erlangen und zu fördern. Auch sprachunspezifische Einflüsse, wie beispielsweise das Dekodieren von Wörtern, können Zweitschriftlernende laut Spruck Wrigley (2008) schneller meistern, da sie bereits in einer Sprache schreiben und lesen können.

In der angewandten Linguistik und der Zweitspracherwerbsforschung wird darauf hingewiesen, dass sprachlich strukturiertes Wissen besonders dann gut transferiert werden kann, wenn Überschneidungen zwischen der Ausgangs- und der Zielsprache gross sind, die Lerner generell einen umfangreichen Wissensstand aufweisen und über ein entsprechendes Repertoire in einer weiteren Sprache verfügen, die idealerweise mit der Ausgangs- und der Zielsprache verwandt ist. Zu bedenken ist dabei allerdings auch: Je besser die Kompetenzen in der Zielsprache sind, desto weniger wird aus der Ausgangssprache transferiert (vgl. Berthele 2018, S. 65–66).

In Bezug auf Zweitschriftlernende stellt sich die Frage, inwiefern ein solcher Transfer stattfindet und inwieweit dieser die Lernprogression des Teilnehmenden unterstützen kann. Auf diese Frage wird in der Diskussion, auf Basis der Befunde dieser Arbeit, näher eingegangen (s. Abschn. 7.3.2).

Die Mehrzahl der Autoren ist sich einig, dass der Prozess des **Schreiberwerbs** in der Zweitsprache langsamer ist als der des Lesens (vgl. Brooks et al. 2001, S. 77–79; Spruck Wrigley 2008, S. 3; Tranza/Sunderland 2009, S. 22), vor allem, da vielfach Interferenzen vorkommen, d. h. Komponenten aus der Erstsprache in die Zweitsprache übernommen werden, die dort so nicht vorkommen. Entsprechend weisen Zweitschriftlernende häufig Schwierigkeiten im Rechtschreiberwerb auf (vgl. Cook 2005, S. 427; Spruck Wrigley 2008, S. 3).

Es gibt eine Reihe von Studien zum Einfluss der Erstsprache (L1) auf die Zweitsprache (L2) im Schreiben. Sie kommen alle zum Schluss, dass Lernende mit geringen Schreibkompetenzen in

der L1 auch entsprechend niedrige Schreibkompetenzen in der L2 aufweisen.⁸ Das bedeutet, dass die Schreibkompetenz in der L2 durch die Schreibkompetenz der L1 vorhergesagt werden kann (vgl. Schoonen et al. 2003, S. 166; Leki et al. 2008, S. 106).

Obwohl Forschungen bezüglich Erst- und Zweitschrifterwerb existieren,⁹ bestehen Forschungslücken im Bereich Zweitschrifterwerb von Flüchtlingen und Migranten mit und ohne Bildungshintergrund, da viele Studien mit Studierenden durchgeführt werden. Das Statement von Gillespie (2001) ist damit nach wie vor aktuell: «To date, we know relatively little about how the development of writing ability in adult literacy learners compares with that of young children or of basic writers at the college level» (ebd., S. 91).

Diese Aussage kann durch die Analyse von Young-Scholten (2018) untermauert werden, welche die Präsentationen der Konferenzen des Vereins ‚Literacy Education and Second Language Learning for Adults‘ (LESLLA) zwischen 2005 und 2017 analysiert hat. Young-Scholten zeigt, dass in diesem Rahmen lediglich 2 % der Präsentationen den Schreiberwerb behandelten. Im Vergleich dazu erreichte der Leseerwerb 7 % und das mündliche Sprachenkönnen 4 %. Das Hörverstehen erscheint nicht auf der Liste, was entweder bedeutet, dass es nie bei einer Konferenz thematisiert wurde oder, dass der Prozentsatz tiefer als .20 ist. Die drei Hauptthemen der Konferenzen zwischen 2005 und 2017 waren verschiedene Fähigkeiten (34 %), Praxiswissen (12 %), gefolgt vom Thema ‚Literacy‘ (11 %) (vgl. ebd., S. 14). Auch anhand der mündlichen Befragung der Konferenzteilnehmenden durch Young-Scholten in Palermo (2018) zeigte sich, dass die *Erstschrift-* und *Zweitschrifterwerbsforschung* mit der Zielgruppe der Analphabeten oder Zweitschriftlernenden nicht zu den fünf wichtigsten Themen zählen, und das, obwohl eines der Ziele von LESLLA die Förderung der Forschung im Bereich des Schriftspracherwerbs ist.

Neben Schreiben ist auch **Lesen** eine Schlüsselkompetenz (vgl. Nold/Willenberg 2007, S. 23), hauptsächlich im Hinblick auf Alphabetisierungskurse, die vielfach die Fertigkeiten Lesen und Schreiben fokussieren. In Alphabetisierungskursen und auch in dieser Forschungsarbeit geht es nicht um die komplexe und vielschichtige Lesekompetenz (vgl. z. B. Rosebrock et al. 2017, S. 13–16), sondern um die basale Lesefähigkeit. Deren Erwerb ist noch vor der eigentlichen Lesekompetenz anzusiedeln, es geht um die beginnende Fähigkeit des Lesens (vgl. Dürscheid

⁸ Für eine detailliertere Diskussion siehe Leki et al. (2008, S. 48–49).

⁹ Für einen Überblick siehe Grabe und Zhang (2016, S. 342–355).

2016, S. 243). Wie bereits in Abschnitt 1.2.1 beschrieben, wird dem Lese- und Schreiberwerb in den zwei grundlegenden Modellen von Frith (1985) und Günther (1986) die jeweils gleiche Bedeutung beigemessen. Beide Fähigkeiten werden als ganzheitliche Prozesse angesehen, die sich gegenseitig bedingen und integrativ erfolgen (vgl. Dürscheid 2016, S. 244).

Obwohl der Lese- und Schreiberwerb miteinander gekoppelt sind, ist das Forschungsinteresse an ersterem deutlich grösser.¹⁰ Der Bereich Zweitspracherwerb mit schriftunerfahrenen Lesenden oder erwachsenen Zweitschriftlernenden mit Flucht- oder Migrationshintergrund ist jedoch ebenfalls nur gering vertreten. Hervorzuheben ist hier die Forschungsarbeit von Feldmeier (2010b), der Leseanfänger im ersten Alphabetisierungskurs (Grundalphabetisierung) untersucht hat. Das Ziel war, zu erfahren, wie Leseanfänger vorgehen, welche Strategien sie anwenden und welchen Schwierigkeiten sie begegnen. Landgraf et al. (2012) haben primäre Analphabeten und Zweitschriftlernende untersucht: Insgesamt nahmen 47 primäre Analphabeten und 41 Personen mit Lese- und Schreibkenntnissen teil, wovon jedoch lediglich drei als Zweitschriftlernende klassifiziert werden können, da die restlichen Personen deutscher Erstsprache waren. Die Autoren haben in der Studie analysiert, ob die Schriftsprachfähigkeiten bei primären Analphabeten anhand der phonologischen Bewusstheit vorhergesagt werden können, und konnten zeigen, dass diese ein stärkerer Prädiktor für die Alphabetisierung ist als beispielsweise die Anzahl der Schuljahre. Zu nennen ist ausserdem die Studie von Kurvers (2015), welche die Entwicklung der Worterkennungsfähigkeit bei sechs primären Analphabeten und sechs Zweitschriftlernenden untersucht und nachgewiesen hat, dass Zweitschriftlernende die geschriebenen Wörter sofort erkannt haben und demnach ein automatisches Dekodieren stattgefunden hat.

Zusammenfassend lässt sich festhalten, dass der Forschungsstand im Schriftspracherwerb, im Lesen und Schreiben, in Bezug auf Zweitschriftlernende vergleichsweise dünn ist.

1.3 Zweitschriftlernende

Trotz fehlender Schriftsprachkenntnisse an ihrem permanenten/vorübergehenden Aufenthaltsort können Zweitschriftlernende nicht als Analphabeten kategorisiert werden, da sie im Gegensatz zu diesen in einer Sprache mindestens funktional alphabetisiert sind (vgl. Hammann et al. 2013, S. 24). Das bedeutet, zumindest im deutschsprachigen Raum, dass

¹⁰ Für einen Forschungsüberblick siehe Grabe und Zhang (2016, S. 342–355).

diese Personen über Schriftsprachkenntnisse verfügen, die den gesellschaftlichen Mindestanforderungen genügen (vgl. Abraham/Linde 2009, S. 92). Was zu diesen Mindestanforderungen gezählt bzw. darunter verstanden wird, ist je nach Gesellschaft unterschiedlich (vgl. Hubertus 1995, S. 252; Panagiotopoulou 2001, S. 60–62; Egloff et al. 2011, S. 13–14; Tröster/Schrader 2016, S. 44–45). Um bestimmen zu können, wie gross die Lese- und Schreibkenntnisse der betroffenen Person sein sollten, müssen die konkreten Gegebenheiten der jeweiligen Gesellschaft genau bekannt sein. Daher ist funktionaler Analphabetismus bis heute nicht an ein allgemeingültiges Niveau gebunden, sondern vielmehr eine vage Kategorie, die sich zudem je nach Land unterscheidet. In der Praxis wird zur Einteilung in diese Gruppe deshalb häufig die Anzahl absolvierter Schuljahre als ausschlaggebendes Kriterium verwendet. Wie Feldmeier (2010a, S. 21–22) jedoch explizit vermerkt, stellt Schule allein noch keinen Garanten für Lernerfolg dar. Da sich die Schulsysteme in den Herkunftsländern zumeist von denen in den Zielsprachenländern unterscheiden, kann es vorkommen, dass eine Person über mehrjährige Schulerfahrung verfügt, diese aber nicht gewinnbringend in den Unterricht einbringen kann (vgl. ebd., S. 20–21). Zudem können die Bildungssysteme der jeweiligen Herkunftsländer nur schwer erfasst werden und Lernen und Lernerfolg sind wesentlich von individuellen Faktoren geprägt. Inwiefern im Bereich von Flüchtlingen somit mit Recht von funktionalem Analphabetismus gesprochen werden kann, ist aus den genannten Gründen nur schwer zu beurteilen.

Damit Zweitschriftlernende in Deutschland einen *Zweitschriftlernerkurs* beginnen dürfen, müssen sie «berufliche Abschlüsse und/oder Schulabschlüsse mit mehrjähriger Schulausbildung aus ihrem Heimatland» aufweisen und

können in mindestens einer Sprache mit nicht-lateinischen Schriftzeichen flüssig lesen und schreiben und sind in dieser soweit alphabetisiert, dass sie in der Lage sind, mit der Schriftlichkeit in allen Lebensbereichen kompetent umzugehen. Somit verfügen sie mindestens über diejenigen schriftsprachlichen Kompetenzen, die minimal erforderlich sind und als selbstverständlich vorausgesetzt werden, um den jeweiligen gesellschaftlichen Anforderungen gerecht zu werden (BAMF 2018a, S. 10).

Zuerst wird im Zweitschriftlernerkonzept aus Deutschland demnach auf die Schulerfahrung Bezug genommen und anschliessend werden die gesellschaftlichen Mindestanforderungen als Kriterien genannt. Im Rahmen der Einstufung der Teilnehmenden, die durch das Goethe-Institut für das BAMF entwickelt wurde (vgl. Perlmann-Balme/Dengler 2007), werden diese mündlich zu ihrer Schulbiographie und eventuellen -abschlüssen sowie zu ihrem Beruf bzw. ihrer Berufsausbildung befragt. Laut Perlmann-Balme und Dengler ist es sehr wahrscheinlich, dass Personen mit weniger als vier Jahren an Schulbildung einen Alphabetisierungskurs

benötigen (vgl. Perlmann-Balme/Dengler 2007, S. 11). Auch wenn die gesellschaftlichen Mindestanforderungen im Deutschen Konzept als Kriterium betrachtet werden, fließen in die abschliessende Beurteilung der Einstufung durch das Goethe-Institut nur die Berufsabschlüsse und die Anzahl der Schuljahre ein.

Im englischsprachigen Raum hingegen werden Adult ‚ESL‘ (English Second Language Learners) mit weniger als zehn absolvierten Schuljahren als Personen mit ‚literacy needs‘ klassifiziert, die Hilfestellungen bei der Entwicklung von Strategien und Fähigkeiten benötigen. Im deutschsprachigen Raum könnten diese als funktionale Analphabeten bezeichnet werden. Personen mit weniger als drei Jahren Schulerfahrung werden als ‚ESL Literacy Learners‘ kategorisiert, denen eine umfangreiche Instruktion, Beratung und Unterstützung zusteht (vgl. Acevedo et al. 2015, S. 5–6) und die im deutschsprachigen Raum in einen regulären Alphabetisierungskurs eingestuft werden würden.

Es liegt durchaus nahe, die Anzahl der Schuljahre als massgebend für die Einstufung als Zweitschriftlernende zu betrachten, denn dieses Kriterium kann relativ einfach erfasst werden. Weitaus umständlicher erscheint es, die tatsächlichen Lese- und Schreibkenntnisse der Personen zu erfassen und durch geschultes Personal beurteilen zu lassen. Hinzu kommt, dass ein entsprechender Test auf Deutsch bis heute nicht existiert. Im Englischen kann das ‚Native Language Literacy Screening Device‘ zurate gezogen werden (vgl. New York State Education Department 1999), das jedoch häufig durch nicht geschulte Personen mit ungenügenden Kenntnissen der zu beurteilenden Zielsprache durchgeführt und bewertet wird.

Im Optimalfall benötigen Zweitschriftlernende laut Feldmeier (2010a) lediglich eine Einführung in das lateinische Schriftsystem (vgl. ebd., S. 20), da sie im Gegensatz zu primären/totalen Analphabeten üblicherweise über Kompetenzen in einer Schriftsprache verfügen: Sie wissen, was ein Wort ist und auch, wie z. B. Wörter synthetisiert werden, und sie erkennen Reime (vgl. ebd.) – gerade dieser Punkt fällt unter die phonologische Bewusstheit und ist für das Erlernen eines alphabetischen Schriftsystems von wesentlicher Bedeutung (s. Abschn. 1.4.2). Zweitschriftlernende sind demnach üblicherweise an das Lernen gewöhnt, verfügen über Schulerfahrung und können verschiedene Lernstrategien, z. B. der Umgang mit einem (digitalen) Wörterbuch, anwenden. Hier stellt sich die Frage, ob eine differenziertere Unterscheidung nach Niveau in der mitgebrachten Sprache angebracht wäre, damit den individuellen Lese- und Schreibkenntnissen der betroffenen Personen Rechnung getragen wird. Zweitschriftlernende mit hohen Lese- und Schreibkenntnissen in einer Sprache bringen

zusätzliches Wissen über die Sprache, zudem üblicherweise mehr Lernstrategien und Lernkompetenzen mit und könnten möglicherweise in einem Einführungskurs eine schnellere Lernprogression als Zweitschriftlernende mit geringeren Lese- und Schreibkenntnissen erreichen, die eine detailliertere Einführung benötigen. Inwiefern eine solche Klassifizierung sinnvoll sein kann, wird in Abschnitt 9.1 näher diskutiert.

1.3.1 Forschungsstand Zweitschriftlernende

Der Begriff Zweitschriftlernende und das dahinterstehende Konzept sind im deutschsprachigen Raum eher unbekannt. Im englischsprachigen Kontext wird es hingegen schon länger diskutiert, nicht zuletzt, weil die Migrationsthematik dort prominenter im öffentlich-medialen und politischen Diskurs verhandelt wird. Dieses öffentliche Interesse ist bisher nur zögerlich in die akademische Diskussion vorgedrungen, weshalb die Forschungslage über Zweitschriftlernende mit Fluchthintergrund nicht sehr umfassend ist; Gleiches gilt für den Analphabetismus von Flüchtlingen und Migranten. Wenn in der Literatur vom Zweitspracherwerb von Flüchtlingen die Rede ist, dann meist, dass diese in diesem Bereich weniger erfolgreich sind, da sie weniger in die Vorbereitung der Migration investiert haben und/oder durch Traumata und Stress beeinträchtigt sind (vgl. Scheible/Rother 2017, S. 7–12). International hat der Bildungsprozess von Migranten seit 2005, mit der Gründung von LESLLA, zunehmend an Beachtung gewonnen. Ziel des Vereins ist es, Forschende verschiedener Wissenschaftsbereiche, Entscheidungsträger und Praktiker verschiedener Herkunftsländer zu vereinen und die Forschung zum Spracherwerb und Schriftspracherwerb von Jugendlichen und Erwachsenen zu fördern (LESLLA o. J.).

Im Folgenden wird in chronologischer Reihenfolge näher auf Forschungen zu Zweitschriftlernenden mit Fluchthintergrund eingegangen.¹¹ Studien zu sogenannten ‚low-literate Learners‘¹² werden nicht berücksichtigt, da bei diesen schwer zu erfassen ist, ob sie in einer Sprache zumindest funktional alphabetisiert sind.

Rother (2010) hat im ‚Working Paper 29‘, das durch das BAMF in Auftrag gegeben worden ist, Teilnehmende aus Alphabetisierungskursen befragt. Das Ziel war die Überprüfung der Wirksamkeit und Nachhaltigkeit der Integrationskurse. Zur Beantwortung dieser Fragen wurden rund 500 ausgefüllte Teilnehmer-Fragebogen und einige Kursleiter-Fragebogen

¹¹ Für einen Forschungsüberblick im Alphabetisierungsbereich ab 2000 siehe Guerrero Calle et al. (im Erscheinen).

¹² Eine relevante Forschung zu ‚low-literate Lerner‘ ist beispielsweise Tarone, Bigelow und Hansen (2009).

herangezogen. Die Resultate zeigen, dass in Alphabetisierungskursen mehrheitlich Frauen (89 %) unterrichten, wobei etwa die Hälfte davon einen Migrationshintergrund aufweist. Die Lehrpersonen können auf einen grossen Erfahrungsschatz im Bereich Deutsch als Zweitsprache (DaZ) zurückblicken, sind aber mehrheitlich unerfahren im Unterrichten von Alphabetisierungskursen. In den Kursen vertreten waren u. a. 21 % Zweitschriftlernende (mind. Niveau B2 und höher in ihrer L1) und 19 % in einem nichtlateinischen Alphabet funktional Alphabetisierte (Niveau A1–B1 in ihrer L1) (vgl. Rother 2010, S. 47–49). Dabei ist jedoch zu beachten, dass die Einteilung in primäre und funktionale Analphabeten und Zweitschriftlernende durch die Lehrpersonen erfolgt ist (vgl. ebd., S. 27). In Anbetracht ihrer geringen Erfahrung im Unterrichten von Alphabetisierungskursen bleibt daher die Frage offen, wie aussagekräftig diese Einstufung der Teilnehmenden ist. Laut der Autorin konnten sie sich zwar hinsichtlich der schriftsprachlichen Kenntnisse der Teilnehmenden auf die Niveaubeschreibung des GER beziehen, doch dieser beinhaltet weder ein spezielles Kapitel über verschiedene Schriftsysteme noch eines zur Einstufung von schriftsprachlichen Kenntnissen in der Erstsprache. Hier lässt sich fragen, inwiefern die Studie aussagekräftig ist, wenn alle Ergebnisse letztendlich auf diesen Einstufungen aufbauen. Rother (2010) hat anhand dieser Differenzierung zudem eine Operationalisierung von primären und funktionalen Analphabeten sowie Zweitschriftlernenden vorgenommen, gemäss der Zweitschriftlernende mindestens ein Niveau B2 in der Erstsprache aufweisen müssen um als solche zu gelten.

Scheithauer (2010) hat in ihrer Studie sieben Personen untersucht, von denen vier Teilnehmende als primäre Analphabeten und drei als funktionale Analphabeten klassifiziert wurden. Die Einteilung erfolgte gemäss Selbsteinschätzung der Teilnehmenden. Ob die funktionalen Analphabeten nicht doch als Zweitschriftlernende hätten eingestuft werden können, bleibt hier offen. Ziel der Studie war es, zu ermitteln, inwieweit die phonologische Bewusstheit Einfluss auf die schriftsprachlichen Kompetenzen hat. Die phonologische Bewusstheit wurde mit einem aus acht, mehrheitlich analytischen Aufgabentypen umfassenden Diagnoseinstrument erfasst. Die schriftsprachlichen Kompetenzen wurden durch «das Legen, Schreiben und Lesen von Silben und Wörtern» (ebd., S. 113) überprüft, ohne dass das Verfahren weiter spezifiziert wird. Die Autorin konnte zeigen, dass die phonologische Bewusstheit (s. Abschn. 1.4.2) einen signifikant positiven Einfluss ($p < .05$) auf die schriftsprachlichen Kompetenzen zeigt.

Schuller et al. (2012) haben im ‚Working Paper 42‘, das im Auftrag des Bundesministeriums des Innern (BMI) entstanden ist, die Entwicklung und Fortschritte der Deutschkenntnisse bei Alphabetisierungskursteilnehmenden evaluiert. Die Hintergrundinformationen und die Operationalisierung dazu wurden durch das oben diskutierte ‚Working Paper 29‘ (Rother 2010) geliefert. Zu Beginn konnte eine Stichprobengrösse von 257 und zum Kursabschluss eine von 263 verzeichnet werden. Dieser Unterschied in der Stichprobengrösse ist dadurch begründet, dass durchgehend neue Teilnehmer die Kurse beginnen konnten. Die Studie zeigt, dass Alphabetisierungskursteilnehmende und somit auch Zweitschriftlernende ihr Sprachniveau im Kursverlauf gemäss Lehrpersoneneinschätzung im Durchschnitt mindestens um ein Level nach GER steigern konnten. Zweitschriftlernende schätzten die Lerngeschwindigkeit im Rahmen der Alphabetisierungskurse jedoch häufig als zu langsam ein. Sie wiesen ein vergleichsweise hohes Bildungsniveau und zum Zeitpunkt der Befragung die kürzeste Aufenthaltsdauer in Deutschland auf. Nach 900 Unterrichtslektionen wurden Zweitschriftlernenden im Vergleich zu den restlichen Teilnehmenden überdurchschnittlich gute Deutschkenntnisse durch die Lehrpersonen attestiert, wobei sie z. T. sogar 100 Lektionen weniger Unterricht hatten als primäre Analphabeten, da sie später in den Kurs einsteigen konnten. Detaillierter betrachtet, zeigen die Ergebnisse bei Zweitschriftlernenden folgende Progression:

- Hören: Zu Kursbeginn lagen 13 % im A2-Bereich und bei Kursende 53 %.
- An Gesprächen teilnehmen: Am Kursanfang befanden sich 15.6 % auf A2-Niveau und zum Schluss 59 %.
- Zusammenhängend sprechen: 16 % Zweitschriftlernende konnten zu Beginn und 53 % am Kursende auf Niveau A2 eingestuft werden.
- Lesen: Am Kursanfang wurden 18 % ins Niveau A1 und zum Kursschluss 84 % auf diesem Niveau eingestuft.
- Schreiben: Zu Beginn befanden sich 16 % auf Niveau A1 und zu Kursende 69 %.

Vor allem im Lesen und Schreiben weist mehr als die Hälfte der Zweitschriftlernenden eine erhebliche Lernprogression auf. Zusammenfassend zeigt die Studie, dass funktionale Analphabeten, die in einem nichtalphabetischen Schriftsystem alphabetisiert sind, zusammen mit Zweitschriftlernenden den stärksten Sprachzuwachs während des Kursverlaufs verzeichneten (vgl. Schuller et al. 2012, S. 5–8, 43–45). Es muss jedoch angemerkt werden, dass die Differenzierung der Kursteilnehmenden in die Analphabetengruppe bzw. zu

Zweitschriftlernenden wie bei der Studie von Rother (2010) durch die Lehrpersonen vorgenommen wurde, weshalb die Ergebnisse ebenfalls unter Vorbehalt zu betrachten sind. Albert et al. (2015) haben im Projekt *Alphamar* untersucht, welche Vorgehensweisen, Materialien und methodischen Ansätze sich besonders für Alphabetisierungskurse eignen. Die Mixed-Methods-Studie, in der mit selbsterstellten Lernfortschrittskontrollen, halboffenen Interviews der Kursteilnehmer, Kursleiterdokumentationen und Fragebogen der Lehrpersonen gearbeitet wurde, weist eine Stichprobengrösse von 38 Teilnehmenden auf. Davon waren 52.6 % primäre, 28.9 % funktionale und 2.6 % sekundäre Analphabeten sowie 15.8 % Zweitschriftlernende.¹³ Die Kursteilnehmenden waren auf insgesamt vier Kursgruppen verteilt. Die Kurseintrittsinterviews der Teilnehmenden wurden mit Hilfe von Dolmetschern in der jeweiligen Erstsprache durchgeführt. Folgende Methoden wurden auf ihre Erfolge hin über den Zeitraum von 360 Stunden untersucht (vgl. ebd.):

1. Die phonetische Methode hatte zum Ziel, einzelne Laute wahrnehmbar zu machen und dadurch das phonologische Bewusstsein aufzubauen.
2. Der methodische Ansatz nach Maria Montessori hat die Sprachfertigkeiten in wesentlich kleinere Schritte unterteilt und arbeitete mit einer Drei-Stufen-Lektion: Darbietung, Wiedererkennen und Produktion.
3. Mit dem Lesen durch Schreiben und sogenannten Buchstabentabellen sollte der Schriftspracherwerb möglichst selbstgesteuert vonstattengehen.
4. Die Silbenmethode arbeitete, wie der Name andeutet, mit Silben.
5. Mit der Morphemmethode wurde erst gearbeitet, nachdem alle Buchstaben eingeführt worden waren. Das Ziel der Methode war die Unterstützung in der Vermittlung von orthografischen oder grammatikalischen Phänomenen.
6. Der Rückgriff auf die Muttersprache wurde punktuell eingesetzt und war demnach keine kurstragende Methode, wie dies bei den vorangehenden Methoden der Fall war. Das Ziel bestand darin, den Zugang zu bestimmten Bereichen zu vereinfachen und zu erleichtern.
7. Das spielerische Lernen wurde auch nur punktuell eingesetzt und sollte den Lernprozess fördern.

¹³ Die addierten Prozentzahlen ergeben nur 99.9 %, was auf die Rundung der Dezimalstellen durch Albert et al. zurückzuführen ist.

In Bezug auf Zweitschriftlernende lässt sich aus der Studie Folgendes herauslesen: Die Methode nach Montessori ($n = 16$) bewirkte den höchsten Lernzuwachs¹⁴ (+.74), gefolgt vom Rückgriff auf die Muttersprache ($n = 23$, +.70) und der Silbenmethode ($n = 12$, +.56). Lediglich das spielerische Lernen hat einen Rückschritt ($n = 25$, -.50) bei der Zielgruppe bewirkt. Es ist jedoch zu beachten, dass auf prüfstatistische Verfahren verzichtet wurde und die Stichprobengrösse recht klein ist (vgl. Albert et al. 2015, S. 116–119). Inwieweit die Teilnehmenden in die vier Gruppen – primäre, funktionale und sekundäre Analphabeten und Zweitschriftlernende – kategorisiert wurden, ist zudem nicht im Detail ersichtlich und wird eher vage beschrieben: «Einige Teilnehmende hatten in der Heimat bereits die Schule in einem normalen Umfang besucht, sodass sie die Schrift ihrer Muttersprache gut beherrschten» (vgl. ebd., S. 14). Sollte die Kategorisierung nur anhand der Schulbildung im Heimatland erfolgt sein, dann kann davon ausgegangen werden, dass die Einstufung der 38 Teilnehmenden nicht akkurat ist (s. Abschn. 1.3).

Die Studie von Kurvers (2015) wurde mit 25 primären erwachsenen Analphabeten, 24 Vorschulkindern und 23 funktionalen erwachsenen Analphabeten durchgeführt. Die letzte Gruppe wird als ‚low-educated‘ bezeichnet und umfasst Teilnehmende, die zumindest grundlegende Lesekenntnisse aufwiesen (vgl. ebd., S. 60–61). Aus der Studie wird nicht ersichtlich, inwiefern es sich bei diesen funktional Alphabetisierten möglicherweise auch um Zweitschriftlernende handelte. Die Forschungsfrage der Studie lautete: «[W]hat knowledge do adult non-literates have about the formal features of their own language compared to young pre-reading children and to low-educated adult readers?» (ebd., S. 60). Die Studie beinhaltete insgesamt elf Instrumente zur Untersuchung der Sprachbewusstheit (metalinguistische Bewusstheit) der Teilnehmenden auf der phonologischen, der Satz- und der Diskursebene. Kurvers (2015) konnte zeigen, dass ‚low-educated‘ Erwachsene sich signifikant von primären Analphabeten unterscheiden und deutlich höhere Ergebnisse erzielten (vgl. ebd., S. 61, 64).

In einer zweiten Studie hat Kurvers (2015) sechs primäre Analphabeten und sechs Zweitschriftlernende untersucht, die hier explizit anhand ihrer Schulbildung als solche klassifiziert werden. Die Forschungsfrage lautete: «how do word-recognition skills develop in

¹⁴ Die sieben Methoden wurden in den vier Untersuchungsgruppen sequenziell und zu verschiedenen Zeitpunkten eingeführt, über einen Zeitraum von je 40 Stunden. Der Lernzuwachs berechnet sich als Differenz zwischen dem Testergebnis der Probanden zu Beginn und am Ende des Einsatzes der Methode. Bei den Lernkontrollen konnten zwischen 0 und 10 Punkten erreicht werden.

adult first-time readers in a second language, and which of the models (if any) of beginning reading best accounts for their development» (Kurvers 2015, S. 66). Dabei hat sich ergeben, dass Zweitschriftlernende im Wortlesen und im Buchstabieren höhere Resultate erzielten als primäre Analphabeten (vgl. ebd., S. 67).

In der dritten Studie hat Kurvers (2015) 322 Erwachsene in der L2 untersucht, wobei ca. 60 % über keine Schulbildung verfügten und ungefähr 35 % mindestens sechs Jahre Schulerfahrung vorweisen konnten. Eines der Ziele war es, herauszufinden, wie viele Stunden die Teilnehmenden für die Stufen A, B und C – die Stufe C ist äquivalent zum Niveau A1 nach GER – benötigen würden. Den Ergebnissen kann allerdings nicht entnommen werden, wie viele Stunden Zweitschriftlernende benötigten, da alle Teilnehmende gemeinsam betrachtet wurden. Die Schulbildung zeigte jedoch eine positive Korrelation mit dem Schreibergebnis. Aus diesem Grund ist anzunehmen, dass Zweitschriftlernende generell höhere Ergebnisse als primäre Analphabeten erzielten. Ungefähr 11 % der Erwachsenen konnten innerhalb von 985 Stunden das Niveau A1 erreichen und einige Teilnehmende bereits nach 300 Stunden. Dabei kann davon ausgegangen werden, dass sich mehrheitlich Zweitschriftlernende in dieser Gruppe befanden (vgl. ebd., S. 70–73).

In den letzten Jahren wurden vermehrt Studien mit Zweitschriftlernenden durchgeführt. Wie aus obiger Darstellung der Forschungsergebnisse ersichtlich, bleibt jedoch oftmals ungeklärt, welche ersprachlichen Kompetenzen die Teilnehmenden in den einzelnen Studien aufweisen. Es ist demnach unklar, ob die Studienteilnehmenden in ihrer L1 primäre oder funktionale Analphabeten oder kompetente Schriftsprachnutzer sind, und somit, ob sie im Sinne der in der vorliegenden Studie benutzten Definition tatsächlich als Zweitschriftlernende gelten.

1.3.2 Die Rolle des Bildungshintergrunds

Wie bereits ausgeführt, wird in der Forschung und der Praxis meist die Anzahl der absolvierten Schuljahre herangezogen, um eine Person als Zweitschriftler zu klassifizieren. So müssen Zweitschriftlernende gemäss dem *Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlerkurs)* (BAMF 2018a) unter anderem über mehrjährige Schulerfahrung und/oder berufliche Abschlüsse verfügen, damit sie einen *Zweitschriftlerkurs* besuchen können (vgl. ebd., S. 10). Was genau unter ‚mehrjährig‘ verstanden wird, ist aus dem Dokument nicht ersichtlich. Wenn versucht wird, dieses Kriterium mit der vom BAMF geförderten Studie von Schuller et al. (2012) zu präzisieren,

dann zeigt sich, dass Zweitschriftlernende einen Mittelwert von 8.0 Schuljahren aufwiesen. Damit liegen sie über den Werten für funktionale ($M = 7.6$) und primäre Analphabeten ($M = 1.7$) (vgl. Schuller 2012, S. 39–40).

Für den englischsprachigen Raum gehen Acevedo et al. (2015) davon aus, dass Zweitschriftlernende mindestens zehn Schuljahre vorweisen müssen, damit keine ‚literacy needs‘ bestehen. Personen mit drei bis zehn Schuljahren hingegen benötigen Unterstützung in der Entwicklung von Fähigkeiten und Strategien und weisen somit ‚literacy needs‘ auf (vgl. ebd., S. 5–6).

In der Schweiz kann weder auf ein Konzept noch ein Rahmencurriculum für Zweitschriftlernende oder Analphabeten zurückgegriffen werden. Im *Rahmencurriculum für die sprachliche Förderung von Migrantinnen und Migranten* (Lenz et al. 2009) geben die Autoren an, dass Personen mit mindestens sechs Jahren Schulerfahrung als schulgewohnt gelten. Da Zweitschriftlernende unter anderem mit schulischen Lerntechniken vertraut sein sollten, könnte dieses Kriterium auch für Zweitschriftlernende herangezogen werden (vgl. ebd., S. 12).

Die empirische Evidenz verdeutlicht, dass der Bildungshintergrund einen positiven Einfluss auf den Zweitschrifterwerb von Flüchtlingen und Migranten hat.¹⁵ Bildung ermöglicht ein effizienteres Lernen im Allgemeinen und von Sprache im Besonderen, da sie Kenntnisse von Lernstrategien und Metawissen impliziert (vgl. Scheible/Rother 2017, S. 16). So zeigt beispielsweise die Studie von van Tubergen (2010), der Flüchtlinge in einer Langzeitstudie hinsichtlich ihres Sprachlernerfolgs im Niederländischen untersucht hat, dass die Schulbildung positiv mit den L2-Fähigkeiten verknüpft ist. Flüchtlinge mit einer höheren Schulbildung zeigten vor allem im Lesen aber auch im Sprechen signifikant bessere Ergebnisse (vgl. ebd., S. 525–526).

Laut Esser (2006b) kann dieser positive Zusammenhang durch unmittelbare und indirekte Wirkungen auf den Spracherwerb erklärt werden. Das heisst, dass Individuen mit ausgedehntem Bildungshintergrund Sprache unmittelbar als eigenen kulturellen Wert ansehen, aber auch den indirekten Nutzen, zum Beispiel den ökonomischen Mehrwert der Sprachbeherrschung, erkennen. Anders ausgedrückt könnte ein möglicher Grund sein, dass Personen mit einer hohen Schulbildung die Sprache an sich anerkennen und wissen, dass sie

¹⁵ Für einen Überblick über die Forschung siehe Esser (2006b, S. 109).

essenziell für den Berufsalltag und ihre professionellen Entfaltungsmöglichkeiten sein kann (vgl. Esser 2006b, S. 109–110).

Inwiefern die Aussage, wonach Teilnehmende mit hohen erschriftsprachlichen Kompetenzen auch im Zweitschrifterwerb höhere Ergebnisse erzielen, zutrifft (s. Abschn. 1.2.3), kann anhand der existierenden Forschungsliteratur nicht beantwortet werden, da keine Forschung mit Analphabeten oder Zweitschriftlernenden mit Fluchthintergrund bzw. mit Migranten vorliegt. Denkbar wäre, dass die Lese- und Schreibkompetenzen in der Erstsprache für den Zweitschrifterwerb gleich wichtig oder gar wichtiger sein könnten als der Bildungshintergrund.

Damit keine voreiligen Schlüsse gezogen werden, sollten die Schreibkompetenzen in der Erstsprache von Zweitschriftlernenden durch erfahrene Prüfer mit hohen Kenntnissen der jeweiligen Sprachen bewertet werden und nicht wie im ‚Native Language Literacy Screening Tool‘ oder beim ‚Einstufungssystem für die Integrationskurse in Deutschland‘ durch die Lehrkraft bzw. die prüfende Person eingeschätzt werden (vgl. Florida Department of Education 2014, 2006, S. 2; Perlmann-Balme/Dengler 2007, S. 17). Dies gilt insbesondere angesichts der Tatsache, dass die Lernertexte durch die Lehrpersonen, die über keine oder wenige Kenntnisse der Ausgangssprache verfügen, nicht verstanden und somit auch nicht schriftsprachlich analysiert werden können. Mit den oben beschriebenen Testinstrumenten können die Lehrpersonen somit lediglich beobachten, wie die Teilnehmenden mit der Aufgabe umgehen, was als Grundlage für eine Einstufung als fraglich gelten muss.

Zusammenfassend lässt sich festhalten, dass die theoretisch begründete und plausible Annahme, der Bildungshintergrund wirke sich positiv auf den Zweitspracherwerb aus, durch empirische Befunde unterstützt werden kann. Es ist anzunehmen, dass dies für Zweitschriftlernende ebenfalls zutrifft, wobei es dazu noch keine Forschung gibt. Da Schulbildung allein keinen Lernerfolg garantieren kann (vgl. Feldmeier 2010a, S. 22), sollte eine Kategorisierung der Teilnehmenden als Zweitschriftlernende nur aufgrund ihrer Schulerfahrung dennoch nur unter Vorbehalt geschehen und möglichst durch ihre Schreibkompetenzen in der Erstsprache ergänzt werden.

1.4 Die Silbe – ein innovatives Lehr-/Lernkonzept?

Bevor die Silbenmethode als didaktisches Mittel im Unterricht mit Analphabeten und Zweitschriftlernenden diskutiert wird, soll die Bedeutung der Silbe im Deutschen erläutert werden – in dem Umfang, der für die Zweitschriftlernerkerse und diese Studie sinnvoll ist.¹⁶

Nach ihrem Schriftsystem gehört die deutsche Sprache zu den Alphabetschriften¹⁷, die silbisch gegliedert und durch Korrespondenzen zwischen Phonemen und Graphemen charakterisiert sind. Diese Phonem-Graphem-Zuordnungen bestehen in alphabetischen Schriften selten aus einer Eins-zu-Eins-Zuordnung (vgl. Nickel 2006, S. 61; Mayer 2016, S. 16). Deutsch wird, wie Buetler et al. (2014) erläutern, wegen seiner transparenten Graphem-Phonem-Korrespondenz zu den relativ transparenten alphabetischen Schriften gezählt: Das heisst, Grapheme und Graphemverbindungen stehen in der Regel für dasselbe Phonem bzw. dieselben Phonemverbindungen. So wird das Graphem ⟨m⟩ im Deutschen meistens durch dasselbe Phonem symbolisiert (vgl. z. B. Perfetti/Dunlap 2008, S. 118; Klicpera et al. 2013, S. 40; Buetler et al. 2014, S. 2; Mayer 2016, S. 17). Andererseits ist die Phonem-Graphem-Zuordnung im Deutschen weniger eindeutig, da einzelne Laute in einem Wort durch unterschiedliche Grapheme symbolisiert werden können. So kann beispielsweise der lange und gespannte Laut [a:] in den Worten *Schal* ⟨a⟩, *mahlen* ⟨ah⟩ oder *Saal* ⟨aa⟩ unterschiedlich verschriftet werden (vgl. ebd., S. 17–18). Dieser Unterschied zwischen der transparenten Graphem-Phonem-Korrespondenz und der weniger eindeutigen Phonem-Graphem-Zuordnung erklärt auch, weshalb das Lesen den Lernenden üblicherweise einfacher als das Schreiben fällt (vgl. Wimmer/Mayringer 2002, S. 272–273).

Die Systematik der deutschen Orthografie lässt sich anhand von drei dominanten Prinzipien erklären (vgl. Mayer 2016, S. 19):

- Phonologisches Prinzip: Stellt die phonologische Struktur der Lautsprache dar.
- Silbisches Prinzip: Beschreibt elementare Zusammenhänge zwischen segmentalen Struktureinheiten im Geschriebenen und Gesprochenen.
- Morphematisches Prinzip: Der Wortstamm muss auch in abgeleiteten und gebeugten Formen erkennbar bleiben.

¹⁶ Die Silbe könnte mit Sicht auf die Phonetik und Phonologie deutlich detaillierter beschrieben werden (s. z. B. Becker 2012, S. 57–68).

¹⁷ Zu den drei grossen Schriftsystemen gehören neben den Alphabetschriften die logografischen Schriftsysteme (z. B. Chinesisch) und die Silbenschriften (z. B. Hiragana) (vgl. z. B. Mayer 2016, S. 13–17).

Insbesondere das silbische Prinzip wird in der Schriftspracherwerbsdidaktik zunehmend thematisiert (vgl. Nickel 2006, S. 61–63) und ist für die deutsche Sprache von grosser Bedeutung: Es regelt die Vokalquantität – die Kürze und Länge der Vokale – und die Vokalqualität – die Ungespanntheit und Gespanntheit der Vokale (vgl. Müller 2010, S. 42).

Die meisten deutschen Wörter enthalten einen zweisilbigen Trochäus, d. h. sie umfassen eine betonte erste Silbe (Hauptsilbe) und eine zweite unbetonte Reduktionssilbe (vgl. Mayer 2016, S. 22). Dies trifft auf 90–95 % aller im Deutschen vorkommenden Wörter zu (vgl. Müller 2010, S. 42–43; Schröder-Lenzen 2013, S. 33); Abweichungen betreffen meist Lehnwörter aus anderen Sprachen oder Funktionswörter (vgl. Mayer 2016, S. 22).

Aus sprachwissenschaftlicher Perspektive ist die Silbe einerseits hilfreich für die Beschreibung von Regularitäten in der deutschen Rechtschreibung, weshalb sie für das Lesen und Schreiben entscheidend ist. Andererseits orientiert sich der natürliche Sprechfluss an der silbischen Struktur von Wörtern, womit die Silbe auch für das Sprechen von zentraler Wichtigkeit ist (vgl. Schröder-Lenzen 2013, S. 33–38; Dahmen/Weth 2018, S. 126–127).

1.4.1 Die Silbenmethode im Unterricht

Die Silbenmethode¹⁸ gehört zur Kategorie der analytisch-synthetischen Methoden und hat die Analyse und die Synthese der Silbe zum Ausgangspunkt. Es werden generell drei Methodenkategorien unterschieden: die synthetischen Methoden, die analytischen Methoden und die analytisch-synthetischen Methoden. Synthetische Methoden führen von kleineren Einheiten (Buchstaben, Laute) zu grösseren Einheiten (Silben, Morpheme, Wörter, Sätze). Die Fähigkeit, unbekannte Wörter zu lesen, steht im Vordergrund und nicht das Lernen von einzelnen Vokabeln. Die analytische Methode geht umgekehrt vor: von grösseren Einheiten zu kleineren. Die analytisch-synthetische Methode versteht sich als Mischform beider Ansätze (vgl. Feldmeier 2010a, S. 53, 65; Rokitzki et al. 2013, S. 99; Feick/Schramm 2016, S. 219). Ziele der Silbenmethode sind die Kombination von gesprochener mit geschriebener Sprache und das zeitgleiche Trainieren von auditiver und visueller Wahrnehmung. Beides ist laut Feick und Schramm (ebd., S. 220) für das Automatisieren des

¹⁸ Die Silbenmethode wurde dem silbenanalytischen Ansatz aus verschiedenen Gründen vorgezogen. Aus eigener Praxiserfahrung mit beiden Methoden kam die Erkenntnis, dass viele Lernende der Zielgruppe deutlich mehr Mühe mit dem silbenanalytischen Ansatz hatten. Zudem ist für diesen Ansatz ein grundlegendes linguistisches Wissen der Kursleitenden notwendig, welches in der Praxis in der Schweiz nicht immer vorausgesetzt werden kann. Zuletzt ist eine der Ausgangslagen der vorliegenden Arbeit die Studie von Albert et al. (2015), welche einen positiven Einfluss der Silbenmethode auf den Zweitspracherwerb nachweisen konnten.

Lese- und Schreibprozesses im Rahmen der Alphabetisierung notwendig. Bei der Silbenmethode wird das sofortige Zusammenziehen von Konsonant und Vokal geübt (vgl. Rokitzki et al. 2013, S. 98–99), was zu einem flüssigeren Lesestil führen kann (vgl. Albert et al. 2015, S. 47).

Die Silbenmethode kann in drei Stufen im Unterricht eingeführt werden. Bevor jeweils mit der nächsthöheren Stufe begonnen werden kann, ist eine sichere Beherrschung der vorherigen erforderlich:

1. In der Elementarstufe wird mit Wörtern ohne Konsonantenhäufungen begonnen (z. B. To-ma-te). Um den Einstieg zu erleichtern, werden zuerst nur lautgetreue Vokale (z. B. ⟨e⟩), Diphthonge (z. B. ⟨ei⟩), Umlaute (z. B. ⟨ue⟩) und Dauerkonsonanten (z. B. ⟨m⟩) behandelt. Anschliessend kommen schwerer hörbare Dauerkonsonanten (z. B. ⟨z⟩) und Stoppkonsonanten (z. B. ⟨d⟩) hinzu. Das Ziel dieser Stufe ist sicheres lautgetreues Schreiben, ausserdem sollte das Lesen einfacher Texte bereits möglich sein.
2. In der Aufbaustufe beginnt die Einführung von Silben mit Konsonantenhäufungen (z. B. Frage). Empfohlen wird, zuerst Konsonantenhäufungen bei Dauerkonsonanten (z. B. ⟨schm⟩) zu behandeln und dann mit Stoppkonsonanten (z. B. ⟨k⟩) fortzufahren.
3. In der folgenden Aufbaustufe, die die sichere Silbengliederung voraussetzt, können Wörter mit bestimmten orthografischen Regelmässigkeiten (z. B. Liebe) und Ausnahmen (z. B. ,Kino') thematisiert werden (vgl. Rokitzki et al. 2013, S. 99–102).

Für den Unterricht mit Kindern wird von Mayer (2018) vor allem die erste Stufe empfohlen, damit die Phonemsynthese und -analyse geübt werden können (vgl. ebd., S. 86–87). Für die Erwachsenenbildung gibt es keine entsprechenden Empfehlungen.

Die auditive und visuelle Wahrnehmung kann u. a. durch rhythmisches Sprechen, Silbenschwingen, Silbenklatschen oder Silbenschreiten trainiert werden. Dadurch erhalten die Lernenden auf körperlicher Ebene ein Gefühl für die Silbensegmentierung. Dies kann dabei helfen, eine deutlichere Artikulation zu erreichen (vgl. Albert et al. 2015, S. 48–49).

Eine häufig anzutreffende Schwierigkeit in Alphabetisierungskursen ist die Skelettschreibung, d. h. das Auslassen der Vokale. Dies kann vor allem bei Persisch und Arabisch sprechenden Personen beobachtet werden, da in diesen beiden Sprachen keine bzw. nur wenige Vokale geschrieben werden. Mit Hilfe der Silbenmethode kann diesen Lernenden bewusst gemacht

werden, dass jede deutsche Silbe einen vokalischen Kern hat, was der Skelettschreibung vorbeugen oder ihr Auftreten zumindest minimieren kann (vgl. Rokitzki et al. 2013, S. 102).

Mit erwachsenen Lernenden hat der Pädagoge Paulo Freire in den sechziger Jahren erstmalig die Silbenmethode für die verarmte brasilianische Bevölkerung benutzt. Dazu wurde ein silbenbasiertes Alphabetisierungskonzept mit der Schlüsselwortmethode¹⁹ kombiniert und eingesetzt. Inwiefern im deutschsprachigen Raum auf dieses Konzept zurückgegriffen werden kann, ist fraglich, u. a. da Konsonantenhäufungen im Portugiesischen weniger vertreten sind als im Deutschen (vgl. Albert et al. 2015, S. 46–47). So kommen diese im Portugiesischen beispielsweise vorwiegend am Silbenanfang vor (praia²⁰) und sind somit weniger komplex als im Deutschen, wo sie am Silbenanfang wie am -ende möglich sind (Strand). In Bezug auf die deutsche Sprache wurde die Silbenmethode insbesondere zwischen 1987 und 1993 im Bereich der Erstlesemethode im erstsprachlichen Grundschulunterricht sowie im Bereich der Legasthenie und des funktionalen Analphabetismus vorangetrieben und ist seither Teil des Unterrichts in diesen Bereichen (vgl. Rokitzki et al. 2013, S. 98).

In den gängigen Alphabetisierungslehrwerken hat die Silbenmethode mittlerweile einen festen Platz (z. B. *Schritte Alpha Plus*, *Alphamar*, *Einfach gut Alpha*) und sie ist auch in der Erwachsenenbildung mit Migranten weit verbreitet (s. z. B. Feldmeier 2010a, S. 63–65; Rokitzki et al. 2013, S. 98–102; Albert et al. 2015, S. 46–52; Feick/Schramm 2016, S. 220).

Die Diskussion zur Silbenmethode ist unüberschaubar geworden, da unterschiedliche Disziplinen beteiligt sind wie Phonetik/Phonologie, Orthografiethorie oder Rechtschreibdidaktik. Als unbestritten gilt aber, dass sich orthografische Regularitäten aus der Silbe ableiten lassen und dass dies nicht nur im Anfängerbereich, sondern auch im späteren Schulalter von Bedeutung ist. Durch die silbische Gliederung erhält der Lernende beim Schreibenlernen eine Unterstützung im Orthografieerwerb. Zudem kommt der Silbe beim Schreibfluss eine wesentliche Rolle zu, da kompetente Schreiber den Schreibfluss an Silbengrenzen unterbrechen (vgl. Werner 2014, S. 113, 115, 120–121). Auch für das Sprechen scheint die Silbengliederung ein geeignetes Mittel zu sein, damit die Sensibilität «auf die Artikulation und auf die phonetisch-phonologischen Strukturen der Sprache [gerichtet ist]»

¹⁹ Bei der Schlüsselwortmethode werden bedeutsame Wörter und Themen durch die Teilnehmenden in den Unterricht eingebracht.

²⁰ Portugiesisch für Strand.

(ebd., S. 117). Da die Silbenmethode besonders bezüglich des Schreibens von Nutzen sein soll, liegt ihre Verbreitung im Unterricht mit Analphabeten und Zweitschriftlernenden nahe.

Ein Nachteil der Silbenmethode ist, dass auf eine Synthese auf Lautebene verzichtet wird. Diese ist aber ab einem gewissen Punkt unerlässlich, da ohne Lautsynthese komplexe Silben (z. B. Deutsch, Mensch) nicht verstanden und beherrscht werden können (vgl. Feldmeier 2010a, S. 63). Insofern scheint die Silbenmethode primär für die ersten Etappen des Schriftspracherwerbs geeignet.

1.4.2 Phonologische Bewusstheit

Phonologische Bewusstheit kann zu den kognitiven Fähigkeiten gezählt werden, die sich mit sprachlichen Struktureinheiten auseinandersetzen. Obwohl die phonologische Bewusstheit bereits empirisch untersucht wurde, ist keine einheitliche Definition vorhanden (vgl. Schnitzler 2008, S. 5). Schnitzler erklärt phonologische Bewusstheit als die Fähigkeit, die lautliche Struktur der gesprochenen Sprache – z. B. Silben, Reime, Laute – zu analysieren und zu manipulieren, ohne auf die Bedeutung einzugehen. Die phonologische Bewusstheit beinhaltet somit «implizites Wissen über den Aufbau von Wörtern aus kleineren Einheiten und den spielerischen Umgang mit Sprache» (Mayer 2016, S. 79).

Laut Feldmeier (2010a, S. 26) fallen folgende Fähigkeiten unter die phonologische Bewusstheit:

- zu hören, ob ein Wort kurz oder lang ist,
- Wort- und Satzakzente zu erkennen,
- Sätze auf Wortebene segmentieren zu können,
- ein lautiertes Wort synthetisieren zu können,
- An-, In- und Auslaute zu erkennen,
- Reime zu erkennen,
- Wörter auf Lautebene segmentieren zu können,
- Lautsubstitutionen innerhalb eines Wortes vornehmen zu können,
- Wörter auf Silbenebene segmentieren zu können.

Wie aus dieser Liste ersichtlich ist, umfasst die phonologische Bewusstheit mehr als ein einziges Konstrukt, weshalb zur Überprüfung auf eine Bandbreite unterschiedlicher Aufgabenstellungen zurückgegriffen werden kann. Vor allem in Förderprogrammen und diagnostischen Verfahren hat die phonologische Bewusstheit einen festen Platz erhalten (vgl. Mayer 2016, S. 79–80).

Die phonologische Bewusstheit spielt gerade bei alphabetischen Schriften eine entscheidende Rolle für den Schriftspracherwerb (vgl. Schnitzler 2008, S. 64), weil damit eine automatisierte Worterkennung und somit die Buchstaben-Laut-Zuordnung und dementsprechend das Lesen unterstützt werden kann (vgl. Martin 2015, S. 4–5). Grund dafür sind Manipulations-,

Synthese- und Analyseprozesse, die zur phonologischen Bewusstheit gezählt werden können und für den Schriftsprachprozess notwendig sind, wobei ein Mangel in der Bewältigung dieser Prozesse in eine sogenannte Lese-Rechtschreibstörung (LRS) münden kann (vgl. Sendlmeier/Oertel 2015, S. 76).

Die phonologische Bewusstheit und mit ihr die Silbenmethode (vgl. Becker 2013, S. 49) spielen im Schriftspracherwerb eine bedeutende Rolle und gelten als dessen Vorläuferfähigkeit (vgl. Feldmeier 2010a, S. 26; Scheithauer et al. 2013, S. 64; Feick/Schramm 2016, S. 218). Hinsichtlich des Aufbaus der phonologischen Bewusstheit wird durch Scheithauer et al. (2013) die Progression von Silben zu Reimen und zum Schluss zu Lauten empfohlen. Gemäss der Autoren sollte zu Beginn des Schriftspracherwerbs der Fokus auf die Arbeit mit Silben gelegt werden, damit die Aussprache und Betonung von Wörtern verbessert und somit neue Sinnzusammenhänge geschaffen werden können. Das Hauptaugenmerk liegt von Anfang an auf der Sensibilisierung für Betonungsmuster im Deutschen, was ein erster bedeutender Aspekt im Hinblick auf die phonologische Bewusstheit ist (vgl. ebd., S. 65–66). Der deutsche Fachdiskurs geht entsprechend davon aus, dass die Silbenmethode vor allem im Anfängerbereich benutzt werden sollte, während im englischsprachigen Bereich dafür plädiert wird, die phonologische Bewusstheit auch im weiteren Kursverlauf zu fördern, um damit die Automatisierung zu sichern (vgl. Davidson/Strucker 2002, S. 313). Dieser Unterschied könnte auf die Sprachen zurückgeführt werden, da Englisch im Gegensatz zum Deutschen eine nichttransparente Sprache ist (vgl. Buetler et al. 2014, S. 2). Es ist aber nicht auszuschliessen, dass eine solche Förderung auch im Deutschen von Bedeutung sein könnte. In vielen Studien konnte ein positiver Zusammenhang zwischen phonologischer Bewusstheit und Schriftspracherwerb bei Kindern nachgewiesen werden (vgl. Scheithauer 2010, S. 121; Mayer 2016, S. 92). So beschreibt beispielsweise Becker (2013) die phonologische Bewusstheit als grundlegend für den erfolgreichen Schriftspracherwerb (vgl. ebd., S. 49). Bei Erwachsenen konnte Scheithauer (2010) in ihrer Forschung zeigen, dass die phonologische Bewusstheit bei erwachsenen Analphabeten – Einstufung als Analphabeten basierend auf der Selbsteinschätzung durch die Teilnehmenden – eine signifikante Korrelation mit schriftsprachlichen Kompetenzen aufweist. Die Studie von Ahlberg (2015) bestätigt zudem, dass erwachsene Analphabeten und Zweitschriftlernende mit der phonologischen Bewusstheit Schwierigkeiten haben, weshalb es sich als sinnvoll erweisen könnte, diese gezielt zu fördern.

1.4.3 Forschungsstand zur Silbenmethode

Die Silbenmethode ist, soweit der Verfasserin dieser Arbeit bekannt, einzig durch das Projekt *Alphamar* bei erwachsenen Analphabeten und Zweitschriftlernenden untersucht worden. Ansonsten kann weder auf nationale noch auf internationale Forschung zurückgegriffen werden.

Im Projekt *Alphamar* wurden neben der Silbenmethode weitere didaktische Methoden bei Analphabeten und Zweitschriftlernenden untersucht. Dabei zeigten Teilnehmende, wenn sie mit der Silbenmethode, aber auch nach der Methode nach Maria Montessori oder unter Rückgriff auf die Muttersprache unterrichtet wurden, einen positiven Lernfortschritt. Die Silbenmethode wies den dritthöchsten durchschnittlichen Lernerfolg im Projekt auf, und bei Lernenden konnte in zehn von zwölf Kompetenzen²¹ ein Lernfortschritt gemessen werden. Lediglich bei der Schreibakkuratheit – Buchstaben korrekt und in erkennbarer Form schreiben (Schriftbild) – und der Vervollständigung struktureller Schemata (z. B. eine Tabelle ausfüllen) wurde ein negativer Lernfortschritt gemessen. Interessant ist, dass die Effektivität der Silbenmethode von der Erstsprache abhängig zu sein scheint: So erreichten Arabisch sprechende Teilnehmende mit dieser Methode den höchsten Lernfortschritt, gefolgt von Persisch sprechenden Teilnehmenden (vgl. Albert et al. 2015, S. 100–117). Aufgrund der eher kleinen Stichprobe von 12 Teilnehmenden und deren unklare Einteilung in die Analphabetengruppe bzw. als Zweitschriftlernender, müssen diese Resultate jedoch unter Vorbehalt betrachtet werden.

²¹ Perzeptive Laut-Buchstaben-Zuordnung, schriftliche Laut-Buchstaben-Zuordnung, Buchstaben-Laut-Zuordnung, Orthografie, semantisches Verständnis, Schreibakkuratheit, Vervollständigen struktureller Schemata, Grammatikanwendung beim gelenkten Schreiben, freies Schreiben, Grammatikanwendung bei mündlichen Übungen, Selbstständigkeit, alltagsbezogene Kommunikationsfähigkeit (vgl. Albert et al. 2015, S. 84–88, 103).

2. Kurskonzepte für Zweitschriftlernende

Wie bereits in der Einleitung thematisiert wurde, existieren in der Schweiz auf bildungspolitischer Ebene weder ein einheitliches Konzept noch entsprechende Alphabetisierungs- bzw. spezielle Kurse für Zweitschriftlernende. Die Trennung von Zweitschriftlernenden und primären Analphabeten ergibt aus unterschiedlichen Gründen Sinn: Zweitschriftlernende wissen beispielsweise, was ein Wort ist oder wie Wörter synthetisiert werden und sind idealerweise an das Lernen gewöhnt (s. Abschn. 1.3). Wegen dieser entscheidenden Vorteile gegenüber primären Analphabeten wurde in dieser Forschungsarbeit das *Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* (BAMF 2018a) aus Deutschland herangezogen, das genau diese Differenzierung zwischen den beiden Gruppen propagiert.

2.1 Zweitschriftlernerkurse Deutschland

Im Jahr 2007 ist das *vorläufige Konzept für einen bundesweiten Integrationskurs mit Alphabetisierung* (vgl. Feldmeier 2007) erschienen, das sich auf das Konzept für Integrationskurse stützt (vgl. Feldmeier 2008, S. 177) und 2015 überarbeitet wurde. Infolge des Anstiegs der Asylbewerberzahlen wurden in Deutschland ab Februar 2017 zudem spezielle Integrationskurse für Zweitschriftlernende konzipiert. Dies geschah aus der Überlegung heraus, dass Zweitschriftlernende unter anderem über grössere Lern- und Stifterfahrungen verfügen, weshalb sie einen früheren Wechsel in einen allgemeinen Integrationskurs erreichen können als primäre Analphabeten, wie ebenfalls im ‚Working Paper‘ von Schuller et al. (2012) erläutert wird. Zweitschriftlernende sollen durch den Kurs in folgender Hinsicht profitieren:

- Die Zweitschriftlernenden erhalten eine bedarfsgerechtere und zielgerichtetere Förderung, um das Zielniveau des Integrationskurses B1 zeitnah zu erreichen. Gleichzeitig wird der Alphabetisierungskurs homogener, was eine stärkere Ausrichtung auf die speziellen Bedarfe der Teilnehmenden ermöglicht.
- Der Zweitschriftlernerkurs knüpft an bereits vorhandene Lese- und Schreibkompetenzen der Teilnehmenden in ihren Erstsprachen an und ermöglicht somit einen effizienteren und schnelleren Sprachlernprozess.
- Die Zweitschriftlernenden können schneller in den Arbeitsmarkt bzw. in die Ausbildung oder in weiterqualifizierte Maßnahmen vermittelt werden (BAMF 2018a, S. 7).

Jeder Zweitschriftlernende erhält zunächst eine Grundförderung, die 600 Kurslektionen à 45 Minuten umfasst: Im Einzelnen sind dies jeweils 300 Kurslektionen für den Basis-Sprachkurs (Abschluss A1) und den Aufbau-Sprachkurs A (Abschluss A2 wird angestrebt). Anschliessend können, unter der Voraussetzung, dass die Kurse bisher regelmässig besucht wurden, weitere

300 Kurslektionen zur Fortsetzung beantragt werden, die als Aufbau-Sprachkurs B verankert sind und mit dem Zielniveau B1 abschliessen sollten. Als regelmässiger Kursbesuch gilt, wenn der Kurserfolg möglich ist, der Lernerfolg nicht wegen Kursabbruch oder Nichtteilnahme am Kurs gefährdet ist und die Person am Abschlusstest ‚Leben in Deutschland‘ teilgenommen hat (vgl. BAMF 2017b, S. 7). Der Kursbesuch endet für jeden Teilnehmenden nach max. 900 Kurslektionen (vgl. BAMF 2018a, S. 8–9). Wie beschrieben, ist das eigentliche Zielniveau nach diesen 900 Kurslektionen B1, wobei das Mindestziel für Alphabetisierungskursteilnehmende das Niveau A2.2 darstellt und für primäre Analphabeten das Mindestniveauniveauziel A2.1 definiert wird. Das A2-Sprachniveau soll hier der funktionalen Alphabetisierung nahekommen (vgl. Feldmeier 2015, S. 12–13).

In der vorliegenden Forschungsarbeit werden lediglich die ersten 300 Kurslektionen untersucht, damit die Stichprobe möglich gross gehalten werden kann. Da Flüchtlinge in der Schweiz eine limitierte Sprachpauschale für Deutschkurse erhalten (s. Abschn. 2.2.2), wäre die Anzahl angemeldeter Teilnehmender bei einer Pilotstudie über 600 oder gar 900 Lektionen wesentlich kleiner gewesen. Zudem entsprechen die ersten 300 Kurslektionen dem geforderten Zielniveau A1 nach dem aktuellen kantonalen Integrationsprogramm (KIP) 2018 bis 2021 des Kantons Bern (s. Abschn. 2.2.1). Im Kanton Bern wird die weitere Sprachbildung individuell behandelt und erfolgt meist parallel zur Arbeit oder zur Ausbildung (vgl. Gesundheits- und Fürsorgedirektion des Kantons Bern 2017, S. 11).

2.1.1 Rahmenbedingungen Zweitschriftlernerkurs

Die Rahmenbedingungen der *Zweitschriftlernerkurse* in Deutschland sind in einem 32-seitigen Dokument beschrieben, wobei auch Verweise auf das *Konzept für einen bundesweiten Integrationskurs* (BAMF 2015) enthalten sind. Im Folgenden wird näher auf diese Rahmenbedingungen eingegangen, da sich die Zweitschriftlernerkurse, die für diese Studie durchgeführt wurden, an diesen Bedingungen orientieren.

2.1.1.1 Kursteilnehmer und Einstufung

Damit ein *Zweitschriftlernerkurs* besucht werden darf, müssen folgende Indikatoren erfüllt sein:

- [Die Teilnehmer] bringen berufliche Abschlüsse und/oder Schulabschlüsse mit mehrjähriger Schulausbildung aus ihren Heimatländern mit.
- Sie können in mindestens einer Sprache mit nicht-lateinischen Schriftzeichen flüssig lesen und schreiben und sind in dieser soweit alphabetisiert, dass sie in der Lage sind, mit der Schriftlichkeit in allen Lebensbereichen kompetent umzugehen. Somit verfügen sie mindestens über diejenigen schriftsprachlichen Kompetenzen, die minimal erforderlich sind und als selbstverständlich vorausgesetzt werden, um den jeweiligen gesellschaftlichen Anforderungen gerecht zu werden.

Dabei gilt zu beachten, dass in Deutschland sowie in anderen Industriestaaten Schriftsprache eine wichtige Bezugsgröße für (fast) alle Lebensbereiche ist. Möglicherweise stammen einige Teilnehmende aus Gesellschaften, in denen die Schriftlichkeit nicht diesen hohen Wert hat, sodass es durchaus sein kann, dass ihre im Heimatland erworbenen und als ausreichend erachteten Schriftkompetenzen für die schriftsprachlichen Anforderungen in der deutschsprachigen Gesellschaft nicht mehr ausreichen. In diesem Fall gelten solche Teilnehmenden als funktionale Analphabeten.

- Sie verfügen über Lese- und Schreibstrategien in ihrer Erstsprache.
- Sie beherrschen beispielsweise verschiedene Lesetechniken, können zwischen narrativen und sachlichen Texten unterscheiden und diese je nach Zweck interpretieren. Sie können beim Schreiben Adressatenbezüge herstellen und Überarbeitungsstrategien anwenden.
- Sie erkennen in ihren (Erst)Sprachen schnell Zusammenhänge in Texten und können je nach Leseintention oder Aufgabenstellung Wesentliches von Unwesentlichem unterscheiden. Sie verfügen möglicherweise über rudimentäre mündliche Kenntnisse der englischen Sprache und beherrschen dadurch einzelne Buchstaben des lateinischen Alphabets.
- Der Umgang mit Lehr- und Lernmedien ist ihnen weitgehend vertraut.
- Der Umgang mit Texten in der deutschen Sprache bereitet ihnen aufgrund fehlender textverarbeitender Routinen im lateinischen Schriftsystem Schwierigkeiten, unter anderem im Hinblick auf Laut-Buchstaben-Zuordnungen, Leerstellen und Satzzeichen und auf die Groß- und Kleinschreibung. Das gilt sowohl für die Rezeption als auch für die Produktion von geschriebenen Texten (BAMF 2018a, S. 10–11).

Diese Merkmale werden im Rahmen des *Einstufungssystems für die Integrationskurse in Deutschland*, das beim jeweiligen Kursträger absolviert wird, erhoben. Es ist lehrwerkunabhängig und basiert auf einer Progression des Schwierigkeitsgrads der Testaufgaben. Es umfasst drei Bausteine: mündliche Sprachkompetenz, schriftliche Sprachkompetenz und Alphabetisierung. Zusätzlich enthält die Handreichung für die Einstufung Indikatoren für einen allfälligen Alphabetisierungsbedarf: Herkunftsland und Erstsprache inkl. Überprüfung der Schreibkompetenz, Schulbesuch und -abschluss – wobei bei weniger als 4 Schuljahren ein Alphabetisierungskurs sinnvoll sein kann (s. Abschn. 1.3) – und Berufsausbildung. Letzter Indikator kann Aufschluss über einen möglichen Analphabetismus geben, beispielsweise kann bei einer im Lehramt tätigen Person Analphabetismus ausgeschlossen werden. Aufgrund dieser Einschätzung wird festgestellt, ob die abzuklärende Person den schriftlichen Teil des Tests oder den Alpha-Baustein auszufüllen hat (vgl. Perlmann-Balme/Dengler 2007, S. 6, 11).

Es gibt zwei Möglichkeiten der Durchführung: Entweder wird direkt mit dem schriftlichen Teil begonnen oder aber es wird zuerst der mündliche Teil absolviert. Im ersten Fall lässt sich ein allfälliger Alphabetisierungsbedarf sofort entdecken. Bevor die definitive Kurszuteilung erfolgt, wird zudem ein Lernberatungsgespräch hinsichtlich der Motivation und der oben aufgeführten Indikatoren geführt (vgl. ebd., S. 10–13).

Ziel des Einstufungsverfahrens ist es, Lese- und Schreibkompetenzen der Betroffenen in der deutschen Sprache zu ermitteln, wobei die mündliche Einstufung der schriftlichen

untergeordnet wird. Die Beurteilung des schriftlichen Teils, d. h. des Alpha-Bausteins, besteht aus einer Analyse der Schreibschrift (Schriftbild) und insgesamt zehn Aufgaben, deren Schwierigkeit zunimmt. Abhängig vom Interviewverlauf mit der abzuklärenden Person kann auf gewisse Aufgaben verzichtet werden, was eine zeitökonomische Durchführung erlaubt. Insgesamt sind im Alphabetisierungskurs neun Module vorhanden: Die Module 1 bis zur ersten Hälfte des Moduls 4 decken das Niveau A1 ab, ab der zweiten Hälfte des Moduls 4 bis Modul 9 wird das Niveau A2 erreicht, wobei Zweitschriftlernende und funktionale Analphabeten üblicherweise in das zweite oder dritte Modul eingestuft werden (vgl. Perlmann-Balme/Dengler 2007, S. 5, 17–18).

2.1.1.2 Kursgefäss

Die maximale Teilnehmerzahl in regulären Alphabetisierungskursen beträgt in Deutschland zwölf Personen (vgl. Feldmeier 2015, S. 30), obwohl nach Erfahrungswerten acht Personen als optimal gelten (vgl. ebd.). Die Teilnehmerzahl in *Zweitschriftlernerkursen* hingegen entspricht der Höchstteilnehmerzahl in allgemeinen Integrationskursen, die bisher unterschiedlich definiert wurde: Im Jahr 2017 lag sie bei 25 Personen (vgl. BAMF 2019a, S. 6), im Folgejahr bei 20 Teilnehmenden (vgl. BAMF 2018a, S. 7) und 2019 bei 23 Lernenden (vgl. BAMF 2019b, S. 7).

Die empfohlenen Wochenunterrichtseinheiten der Alphabetisierungskurse liegen zwischen 12 und 16 (vgl. Feldmeier 2015, S. 15), wobei eine Unterrichtseinheit 45 Minuten dauert (vgl. BAMF 2017c, S. 2). Von einer geringeren Anzahl an Wochenstunden wird aufgrund methodisch-didaktischer Überlegungen abgeraten, da sonst die Gefahr besteht, die erlernten Inhalte wieder zu vergessen (vgl. Feldmeier 2015, S. 144–146). Maximal sind in Alphabetisierungskursen pro Woche 25 Unterrichtseinheiten à 45 Minuten vorgesehen (vgl. ebd., S. 15). In Deutschland werden die *Zweitschriftlernerkurse* üblicherweise in Fünferblöcken durchgeführt, d. h. in fünf Unterrichtseinheiten pro Tag, was 25 Lektionen pro Woche entspricht.

Die Auswahl des kurstragenden Lehrwerks ist den Kursträgern bzw. den Lehrpersonen überlassen, solange es sich um geeignete Unterrichtsmaterialien für Zweitschriftlernende handelt. Dabei kann aus einer Liste für zugelassene Lehrwerke ausgewählt werden, die für die ersten 100 Kurslektionen 23 Lehrwerke (Stand: Oktober 2018) enthält (vgl. BAMF 2018b, S. 5–6). Keines dieser Lehrwerke ist speziell für die Schweiz konzipiert, d. h. die Gespräche und

Themen entsprechen den deutschen Gegebenheiten. Dennoch werden zahlreiche dieser Lehrwerke auch in der Schweiz benutzt.

2.1.2 Inhalte und Ziele des Basis-Sprachkurses

Der Basis-Sprachkurs mit 300 Kurslektionen umfasst drei Abschnitte: Einen ersten Teil zur Einführung in das lateinische Schriftsystem und die Teile zwei und drei für den Sprachkurs mit Zielniveau A1 (vgl. BAMF 2018a, S. 15).

2.1.2.1 Erster Kursabschnitt

Der erste Abschnitt führt in das lateinische Schriftsystem ein und soll das Fundament für die weiteren Kursabschnitte legen. Behandelt werden dementsprechend folgende Inhalte (vgl. ebd., S. 20):

1. Alle 29 Buchstaben des deutschen Alphabets (inkl. Umlaute) sowie Buchstabengruppen, die Kombinationen aus mind. zwei Buchstaben, z. B. ‚ei‘ beinhalten.
2. Häufige Buchstabenkombinationen, die keinen eigenen Lautwert haben, z. B. ‚tz‘.
3. Die Synthesefähigkeit, damit Buchstaben/Buchstabengruppen zu einem lautlichen Ganzen verschmolzen werden können.
4. Die Analysefähigkeit, um gesprochene/gehörte Wörter in phonemische Bestandteile zerlegen zu können.

Das übergeordnete Ziel des ersten Abschnitts besteht darin, die Teilnehmenden dazu zu befähigen, den Sprachkurs erfolgreich fortzusetzen. Dieses Ziel kann in zwei Unterziele aufgeteilt werden: Die Teilnehmenden sollen mit allen im Deutschen verwendeten Schriftzeichen vertraut sein, d. h. die Identifizierung sowie Produktion der Schriftzeichen sollte für sie kein Problem darstellen, und sie sollen die gängigsten Redemittel und Vokabeln kennen, was die Teilnahme am Sprachkurs erleichtern soll. Um das erste Ziel zu erreichen, wurden folgende schriftbezogenen Feinlernziele definiert:

Die Teilnehmenden beherrschen

- alle lateinischen Groß- und Kleinbuchstaben des deutschen Alphabets rezeptiv und produktiv.
- gängige Buchstabenkombinationen des Deutschen rezeptiv und produktiv.
- die wichtigsten orthografischen Muster in der deutschen Sprache.
- die Versprachlichung aller Zahlen im Zahlenraum von 0 bis 1 Million.

Die Teilnehmenden können

- alle Buchstaben des deutschen Alphabets in Druckform und in verständlich geschriebenen Handschriften erlesen.
- verschiedene Silbenstrukturen des Deutschen erkennen und unterscheiden (BAMF 2018a, S. 16).

Zusammengefasst sollen die Teilnehmenden im schriftlichen Bereich die Analyse- und Synthesefähigkeit im Deutschen auf Wort-, Morphem-, Silben- und Laut-/Phonemebene entwickeln (vgl. BAMF 2018a, S. 16).

Hinsichtlich allgemeinsprachlicher Kompetenzen sollen die Lernenden über Folgendes verfügen:

- wichtige bedarfsorientierte Vokabeln und Redewendungen in Form von Mehrworteinheiten (Chunks) für die unmittelbare Alltagskommunikation verstehen und anwenden,
- sich vorstellen und verschiedene Begrüßungsformeln anwenden,
- einfache Anweisungen und Aufgabenstellungen im Kurs verstehen (BAMF 2018a, S. 17).

In den ersten Kurslektionen wird explizit von einer Sprachprogression abgesehen. Sprachliche Inhalte sollen nur so weit thematisiert werden, wie sie den Bedürfnissen der Kursteilnehmenden entsprechen. Nach Abschluss der ersten 100 Kurslektionen wird ein Zwischentest zur Überprüfung der Zielerreichung empfohlen, der jedoch durch die Lehrperson selbst konzipiert werden muss, da keine standardisierten Zwischentests vorliegen (vgl. ebd., S. 17, 21). Dieses Fehlen eines geeigneten Beurteilungsinstruments stellt ein Manko dar, denn dadurch ist keine einheitliche Sprachstandüberprüfung möglich. Gleichzeitig ist sie kennzeichnend für die Forschungslage auf dem Gebiet, da bisher zu wenig empirische Evidenz vorhanden ist, die für die Konstruktion eines solchen Messinstruments nötig wäre.

2.1.2.2 Zweiter und dritter Kursabschnitt

Die beiden folgenden, weiteren Kursabschnitte sollen das Fundament für eine «wachsende sprachliche Selbstständigkeit der Teilnehmenden» (ebd., S. 20) legen. Die dafür nötigen sprachlichen Ressourcen sind: Wortschatz, Grammatik sowie entsprechende Themen, Situationen und Handlungsmuster. Thematisch werden die wichtigsten alltäglichen Lebenssituationen der Teilnehmenden abgedeckt, wie Beruf und Arbeit, Wohnen, Gesundheit und Krankheit (vgl. ebd., S. 20–21).

In diesen beiden Abschnitten sollen die Teilnehmenden ihre schriftbezogenen Kompetenzen ausbauen und festigen, damit ein zunehmend routinierter Umgang mit der lateinischen Schrift möglich wird. Nach Abschluss der insgesamt 300 Kurslektionen sollen die Teilnehmenden mindestens das Niveau A1 erreicht haben, wobei für einzelne Teilnehmende auch ein höheres Kursniveau möglich ist. Das schriftbezogene Kursziel besteht darin, die Lernenden zu ermächtigen, weitgehend routiniert mit der lateinischen Schrift umzugehen. Diese Ziele unterscheiden sich vom GER insofern, als Rücksicht auf ein potenziell langsames Lese- und

Schreibtempo genommen werden muss und die Teilnehmenden folgende Kompetenzen, nebst den Zielvorgaben des GER, erreicht haben sollen (vgl. BAMF 2018a, S. 19):

- in einem angemessenen Tempo von links nach rechts schreiben,
- auf einer Zeile schreiben,
- lautgetreue Vokabeln verständlich und lesbar schreiben (Laut-Buchstaben-Zuordnungen),
- Wörter mit erkennbaren Leerzeichen voneinander abtrennen,
- sowohl die Groß- und Kleinschreibung als auch die Grundregeln der Satzzeichen im Großen und Ganzen aktiv anwenden (BAMF 2018a, S. 17).

Am Ende des dritten Abschnitts, nach 300 Kurslektionen, kann entweder die telc-Prüfung A1 oder das Goethe-Zertifikat A1 durchgeführt werden (vgl. ebd., S. 21).

2.1.3 Didaktik im Basis-Sprachkurs

Die Didaktik soll andragogischen Prinzipien entsprechen, wobei die Grundlagen des Zweitschrifterwerbs zu berücksichtigen sind. In diesem Sinne richtet sich die Methodenauswahl nach den Lernzielen und -inhalten sowie den Voraussetzungen der Zielgruppe – das sind Vorkenntnisse, soziokulturelle Faktoren, Lernhaltung und weitere Aspekte. Der Unterricht soll dabei grundlegend handlungsorientiert, an die kommunikativen Bedürfnisse der Zielgruppe angepasst und interkulturell konzipiert sein (vgl. ebd., S. 14).

Der entscheidende Unterschied zwischen einem regulären Alphabetisierungskurs und dem ersten Abschnitt des *Zweitschriftlernerkurses* besteht darin, dass in letzterem die Einführung in das lateinische Schriftsystem im Vordergrund steht. In üblichen Alphabetisierungskursen hingegen findet eine Art Verschmelzung von Alphabetisierung und dem Erwerb sprachlicher Kompetenzen statt; im *Zweitschriftlernerkurs* bleibt der Erwerb sprachlicher Kompetenzen zunächst im Hintergrund (vgl. ebd., S. 13).

Für den ersten Kursabschnitt werden zwei grundlegende Methoden genannt: die synthetischen Methoden für die Buchstabeneinführung und die analytische Methode zur Hilfe bei der Einführung von alltagsrelevanten Vokabeln. Analytisch-synthetische Methoden, wie die Silbenmethode, werden nicht explizit aufgeführt, sind aber mit den grundsätzlichen Zielen des Unterrichtskonzepts vereinbar. Die Vermittlung von Grammatik und Wortschatz in Form von Chunks erfolgt in diesem Abschnitt nur, wenn es für die Sprachhandlung erforderlich ist und den Alltag der Lernenden unmittelbar betrifft (vgl. ebd., S. 13–14).

2.1.4 Qualifikation der Lehrpersonen

Damit Lehrpersonen in einem *Zweitschriftlernerkurs* unterrichten dürfen, wird eine Zulassung für Integrationskurse benötigt. Diese verlangt entweder eine entsprechende Ausbildung oder den Abschluss eines Qualifizierungslehrgangs, der vom BAMF vorgegeben wird. Das

Bundesamt will dadurch die geeigneten andragogischen und interkulturellen Kompetenzen und eine hohe fachliche Qualifikation der Lehrkräfte in Integrationskursen sicherstellen (vgl. BAMF 2018a, S. 28; BAMF 2018c). Keine Zusatzqualifizierung benötigen Lehrpersonen, die in Deutschland einen Hochschulabschluss in DaF/DaZ erworben haben oder über eine Lehrbefähigung für Deutsch oder eine moderne Fremdsprache verfügen (inkl. Grundschullehramt) (vgl. BAMF 2018d).

Eine verkürzte Qualifizierung von 70 Kurslektionen müssen Personen absolvieren, die entweder über ein anerkanntes (Hochschul-)Zertifikat DaF/DaZ, einen Hochschulabschluss in Germanistik bzw. anderen Neuphilologien oder über andere anerkannte DaF/DaZ-Zertifikate verfügen, aber keinen Unterrichtsnachweis vorlegen können. Die unverkürzte Qualifizierung von 140 Kurslektionen gilt für Personen, die einen Hochschulabschluss in Pädagogik, Sozialpädagogik, Sonderpädagogik, Erwachsenenbildung, Erziehungswissenschaften, Psychologie und andere DaF/DaZ-Zertifikate mit mindestens 100 Kurslektionen aufweisen. Das betrifft auch andere Hochschulabschlüsse oder sprachliche Berufsabschlüsse ohne Hochschuldiplom. Unterrichtserfahrung von über 500 Lektionen kann angerechnet werden und zu einer Verkürzung oder zum Wegfall der benötigten Zusatzqualifizierung führen (vgl. BAMF 2018d).

Eindeutige Regelungen zu den erforderlichen Vorbildungen für potenzielle Lehrpersonen in diesem Bereich hat das BAMF somit bereits eingeführt sowie Zusatzqualifizierungen für Lehrpersonen in diesem Themengebiet geschaffen. Die Finanzierung dieser Qualifizierungsmassnahmen müssen potenzielle Lehrpersonen nicht selbst tragen, sondern es besteht die Möglichkeit, beim Bundesamt eine Rückerstattung anzufordern (vgl. BAMF 2019).

2.2 Kurse in der Deutschschweiz

Für die vorliegende Studie wurden Zweitschriftlernererkurse im deutschsprachigen Teil des Kantons Bern durchgeführt. Um diese Kurse einzuordnen und die Unterschiede zur Situation von Zweitschriftlernenden im derzeitigen Berner Bildungssystem (Stand 2019) herauszuarbeiten, soll an dieser Stelle das aktuelle Integrations- und Sprachförderungssystem des Kantons Bern skizziert werden. Es gilt zu beachten, dass sich die Situation aufgrund der föderalistischen Integrations- und Bildungspolitik von Kanton zu Kanton unterscheidet.

2.2.1 Kantonales Rahmenkonzept in Bern

Analog zur gesamtschweizerischen Entwicklung in der Ausländerpolitik (s. Einleitung), etablierte sich die Idee, dass diese nicht nur die Kontrolle von Niederlassung und Aufenthalt,

sondern auch die Integration regeln soll, auf kantonaler Ebene erst relativ spät. Den Anfang machte die Integrationspolitik im Kanton Bern, als 1996 ein Postulat der Grossrätin Barbara Gurtner im Regierungsrat besprochen wurde. In der Folge wurden zwei Expertenberichte verfasst, die die Situation der Migrationsbevölkerung im Kanton darlegten und Möglichkeiten aufzeigten, wie eine Integrationspolitik auf kantonaler Ebene aussehen könnte. Im Jahr 2003 wurde die Stelle für Integrationsfragen in der Gesundheits- und Fürsorgedirektion (GEF) geschaffen, die seit 2006 als kantonale Fachstelle Integration und als Koordinationsstelle für verschiedene integrationspolitische Akteure tätig ist (vgl. Bader et al. 2013, S. 17–19). Hintergrund dieser Neustrukturierung war die Übergabe der Verantwortung für die Integrationsförderung vom Bund an die Kantone. Dazu entwickelten die Kantone sogenannte kantonale Integrationsprogramme (KIP), die jeweils eine Dauer von vier Jahren umfassen und die ergänzenden Massnahmen der spezifischen Integrationsförderung formulieren sollen. Anhand dieser KIP sollen Ziele und Massnahmen bestimmt und die Umsetzungsorganisation, die Finanzierung und die Qualitätssicherung dazu aufgezeigt werden (vgl. Zuppinger et al. 2013, S. 10). Zu dieser spezifischen Integrationsförderung gehören u. a. die Sprache und Bildung für Ausländer mit Ausweisen C, B, F oder L.²²

Zur Zielgruppe der spezifischen Integrationsförderung gehören somit Personen mit B- und F-Ausweisen, was die Teilnehmenden am Dissertationsprojekt einschliesst. Das erste KIP dauerte von Anfang 2014 bis Ende 2017, d. h. zum aktuellen Zeitpunkt läuft das zweite KIP seit Anfang 2018 bis Ende 2021 (vgl. ebd., S. 10–11). Eines der Ziele besteht darin, dass ab 2020 anerkannte Flüchtlinge innerhalb von sechs Monaten mindestens ein Sprachniveau A1 erreichen und nachweisen können (vgl. Gesundheits- und Fürsorgedirektion des Kantons Bern 2017, S. 30). Die Erreichung dieses Zielniveaus «hat einen direkten Einfluss auf die Höhe der Sozialhilfe», die um bis zu 30 % gekürzt werden kann (Gesundheits- und Fürsorgedirektion GEF und Polizei und Militärdirektion POM 2017, S. 21), sowie auf die Wohnungsumgebung, denn eine eigene Wohnung darf u. a. erst mit einem Sprachniveau A1 bezogen werden (vgl. ebd., S. 19).

²² Vorläufig aufgenommene Flüchtlinge (Ausweis F) erhalten einen für max. 12 Monate gültigen Ausweis und können nach fünf Jahren Aufenthalt in der Schweiz einen Ausweis B beantragen. Flüchtlinge mit Ausweis B müssen diesen nach einem Jahr verlängern und können nach fünf Jahren erfolgreicher Integration oder zehn Jahren eine unbefristete Bewilligung (Ausweis C) beantragen. Kurzaufenthalter (Ausweis L) halten sich befristet üblicherweise unter einem Jahr in der Schweiz auf (vgl. Staatssekretariat für Migration SEM 2011, 2019b)

2.2.2 Vergleich zu Deutschland

Die Alphabetisierungskurse in der deutschsprachigen Schweiz nehmen keine Trennung zwischen Analphabeten und Zweitschriftlernenden vor. Dieses Vorgehen ist, wie bereits ausgeführt wurde, problematisch: Primäre Analphabeten werden meist über- und Zweitschriftlernende unterfordert oder beides passiert gleichzeitig (vgl. Spruck Wrigley 2008, S. 1; BAMF 2018a, S. 7), weshalb es angeraten ist, die Gruppen getrennt zu unterrichten.

Im Kanton Bern gibt es unterschiedliche Anbieter, die Alphabetisierungskurse durchführen. In allen werden Analphabeten und Zweitschriftlernende in gemeinsamen Kursen unterrichtet. Die Grösse der Kursgruppen unterscheidet sich je nach Anbieter und umfasst zwischen acht und 16 Personen. Es gibt in der Regel drei Alphabetisierungsstufen, wobei die erste häufig lediglich von primären Analphabeten besucht wird, die über keinerlei Kenntnisse der Schriftsprache verfügen. Üblicherweise starten Zweitschriftlernende im Alphabetisierungskurs mit den Stufen zwei oder drei. Ein Kurslevel kann zwischen drei und sechs Monaten dauern, je nach Kursfrequenz. Es kommt zudem nicht selten vor, dass Teilnehmende den Kurs wiederholen müssen.

Wieso es problematisch ist, Zweitschriftler zusammen mit primären Analphabeten zu unterrichten, lässt sich anhand eines Beispiels verdeutlichen, das die finanziellen Aspekte beleuchtet: Ein Zweitschriftler wird in einen Alphabetisierungskurs auf Stufe zwei eingeteilt. Sein Kurs dauert sechs Monate bei sechs Stunden pro Woche und kostet 1'000 Franken. Im Anschluss kann er die dritte Alphabetisierungsstufe besuchen, wobei dieser Kurs 1'300 Franken kostet und acht Stunden pro Woche umfasst und ebenfalls sechs Monate dauert. Nach einem Jahr kann der Lernende dann mit einem A1-Kurs beginnen, der ein weiteres Jahr dauert und viermal wöchentlich jeweils zwei Stunden in Anspruch nimmt. Dieser Kurs würde 2'600 Franken kosten, weshalb der Teilnehmer ihn nicht abschliessen darf, da seine Sprachpauschale lediglich 4'300 Franken²³ umfasst und somit nur noch 2'000 Franken übrig sind. Er kann somit nur einen A1.1 Kurs besuchen und muss anschliessend auf freiwillige Deutschkurseangebote zurückgreifen – oder den Unterricht aus der eigenen Tasche bezahlen –, da seine Sprachpauschale ausgeschöpft ist.

Dieses Vorgehen schadet Zweitschriftlernenden auf zwei Ebenen: Ihre Lernprogression kann

²³ Die Sprachpauschale umfasst durchschnittlich 4'300 und soll auf 3'100 gesenkt werden (vgl. Gesundheits- und Fürsorgedirektion GEF und Polizei und Militärdirektion POM 2017, S. 29). Sobald die Pauschale ausgeschöpft ist, kann in gewissen Fällen eine Erhöhung des Kursgeldes beantragt werden. Falls dies nicht möglich ist, dürfen keine weiteren Deutschkurse besucht werden, da die Finanzierung ausbleibt.

durch den verlängerten Alphabetisierungsprozess beeinträchtigt werden (vgl. Schuller et al. 2011, S. 7) und ihre Sprachpauschale wird ineffizient eingesetzt.

Der Unterschied zwischen dem deutschen *Zweitschriftlernerkurs* und dem Status quo in Bern ist offensichtlich: In Deutschland können Zweitschriftlernende in einem für sie angepassten Kurs ihrem Lerntempo folgen und demnach eine raschere Lernprogression erreichen. Das Spracherwerbssystem ist zielgerichteter und bedarfsgerechter, wie dies auch proklamiert wird (vgl. BAMF 2018a, S. 7). Unter Einbeziehung von Grundbildung und Literalität in regulären A1-Kursen in der Schweiz hätte der Kursteilnehmer aus dem Beispiel lediglich den dritten Alphabetisierungskurs besuchen müssen, da er seine Lese- und Schreibkenntnisse im A1-Kurs noch weiter hätte vertiefen können. Damit hätte seine Sprachpauschale für den Abschluss des A1-Niveaus gereicht und es wären sogar noch 400 Franken übrig geblieben, die für weitere Sprachfördermassnahmen hätten eingesetzt werden können.

Teil II: Methodologie

3. Forschungsgegenstand

Dieser Forschungsarbeit liegen zwei Zielsetzungen zugrunde: Zum einen soll die Umsetzbarkeit des *Konzepts für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* (BAMF 2018a, s. Abschn. 2.1) in der Schweiz überprüft werden und zum anderen soll die Wirksamkeit der Silbenmethode (s. Abschn. 1.4) untersucht werden. Zu diesem Zweck wurde eine Interventionsstudie im natürlichen Kursumfeld (Feldexperiment) durchgeführt, die die Effektivität der Treatments – das *Zweitschriftlernerkonzept* und die Silbenmethode – analysiert. Betrachtet wird somit das Verhältnis zwischen Ursache (*Zweitschriftlernerkonzept* und Silbenmethode) und Wirkung (Test- und Lernerfolge). Um diese Kausalität nachweisen zu können,²⁴ wird mit einem experimentellen Design gearbeitet, «[d]enn nur durch Experimente kann man der Frage nachgehen, ob eine kausale Beziehung zwischen zwei (oder mehreren) Umständen oder Phänomenen besteht» (Marx 2012, S. 53). Das Forschungsdesign wurde so gewählt, um den kausalen Effekt der zwei Variablen – *Zweitschriftlernerkonzept* und Silbenmethode – zu untersuchen und die zuvor aufgestellten Hypothesen zu überprüfen (s. Abschn. 3.6). Dies wiederum erlaubt anschliessend eine vorsichtige Generalisierung der Hypothesen in Bezug auf Zweitschriftlernende mit ähnlichen Hintergründen (vgl. ebd.). Das *Zweitschriftlernerkonzept* und die Silbenmethode sind hier die unabhängigen Variablen (UV) und die sprachliche Kompetenz, gemessen als Gesamtlernerfolg oder dem Abschneiden beim telc-Abschlusstest, die abhängige Variable (AV).

Damit die Schlussfolgerungen zu den AV auf die UV zurückgeführt werden können, sind drei Voraussetzungen zu berücksichtigen (Darsow/Felbrich 2014, S. 230):

1. Die UV und die AV müssen kovariieren, d. h., die beiden Merkmale ändern sich in Abhängigkeit voneinander. [...]
2. [Es] muss sichergestellt sein, dass die UV der AV bzw. der beobachteten oder gemessenen Veränderung der AV vorausgeht: UV --> AV. [...].
3. [Es] muss überzeugend festgestellt werden, dass die beobachtete Wirkung mit einer hohen Wahrscheinlichkeit auf das *Treatment* zurückgeführt werden kann und alternative Erklärungsansätze unplausibel sind. [...].

²⁴ Mit statistischen Methoden lassen sich streng genommen keine Kausalitäten beweisen, sondern lediglich Zusammenhänge zwischen zwei Ereignissen. Die Kausalität ergibt sich erst unter Zuhilfenahme zusätzlicher, nicht statistischer Informationen. Dadurch, dass im vorliegenden Fall die UV der AV zeitlich vorausgeht und eine Vielzahl potenziell intervenierender Variablen kontrolliert werden, scheint die Kausalitätsannahme aber zumindest plausibel (vgl. Weiss 2018).

Für die in dieser Studie verwendeten abhängigen Variablen Gesamtlernerfolg und Resultat bei der telc-Abschlussprüfung und die unabhängige Variable Silbenmethode sind alle drei Voraussetzungen gegeben. Wie im statistischen Teil der Arbeit ersichtlich wird (s. Abschn. 6.2.2 und 6.3.2), kovariieren diese UV und AV. Das heisst der Gesamtlernerfolg der Studienteilnehmenden und ihr Prüfungsergebnis ändern sich in Abhängigkeit von der Variable Silbenmethode.

Auch Voraussetzung zwei, dass die UV den AV vorausgehen, ist erfüllt. Die Silbenmethode wurde zu Beginn der Studie eingeführt und durchgehend verwendet, wohingegen die Messung der AV, des Gesamtlernerfolgs und des Abschneidens beim telc-Test, erst nach 200 Lektionen resp. am Schluss der Studie erfolgte. Das heisst es kann ausgeschlossen werden, dass die beiden abhängigen Variablen die Einteilung in das Treatment Silbenmethode beeinflusst haben.

Um die dritte Voraussetzung zu erfüllen, wurde versucht, die Einflussfaktoren so gering wie möglich zu halten. Hierzu wurden möglichst viele theoretisch und auf Erfahrungswerten begründete Störvariablen operationalisiert (s. Abschn. 5.3.4), um den Einfluss intervenierender Variablen möglichst gering zu halten. Anders als z. B. bei Laborexperimenten kann es bei Feldforschungen jedoch schwierig sein, alle unabhängigen Variablen zu definieren und zu kontrollieren. Dafür hat die Feldforschung den Vorteil, dass der zu untersuchende Faktor *Zweitschriftlernernde* nicht aus dem natürlichen Kontext gerissen und somit die externe Validität erhöht wird (vgl. Marx 2012, S. 63).

Für die unabhängige Variable *Zweitschriftlernerkonzept* sind, analog zur Silbenmethode, Voraussetzungen zwei und drei gegeben. Der ersten Voraussetzung kann jedoch streng genommen in der vorliegenden Studie nicht entsprochen werden. In der Schweiz existiert kein vergleichbarer Kurs (s. Abschn. 2.2), weshalb sich keine Varianz in dieser UV herstellen lässt, auch nicht in der simplen binären Form Verwendung/Nichtverwendung *Zweitschriftlernerkonzept*. Damit stellt sich auch keine Kovarianz zwischen dieser UV und den beiden AV ein. Die UV *Zweitschriftlernerkonzept* dient somit nicht der Überprüfung, wie sich dieses Konzept von einem anderen Konzept unterscheidet, sondern inwieweit die Niveauerreichung A1 in ebendiesem Zweitschriftlernerkonzept erreichbar ist. Die UV *Zweitschriftlernerkonzept* wird aus diesem Grund zwar im Rahmen der uni- und bivariaten deskriptiven Statistik berücksichtigt, nicht aber in die linearen Regressionen in den Abschnitten 6.2.3 und 6.3.4 miteinbezogen.

3.1 Experimental- und Kontrollgruppe

Damit eine Wirkung auf ein Treatment zurückgeführt werden kann, werden mindestens zwei Gruppen benötigt, die untersucht werden (vgl. Darsow/Felbrich 2014, S. 231). Üblicherweise wird bei dieser Art von Experimenten mit einer Experimental- und einer Kontrollgruppe gearbeitet. Da Zweitschriftlernende in der Schweiz jedoch bis anhin gemeinsam mit Analphabeten unterrichtet werden und somit noch keine reguläre Unterrichts- oder Lernsituation für diese Zielgruppe existiert, konnte keine Kontrollgruppe im engeren Sinne hinzugezogen werden. In dieser Forschungsarbeit wird deshalb mit den Treatmentgruppen 1 und 2 gearbeitet, wobei für beide die unabhängige Variable des *Konzepts für einen bundesweiten Integrationskurs für Zweitschriftlernende* gilt. Die Treatmentgruppe 1 dient zudem als Kontrollgruppe, da lediglich Treatmentgruppe 2 mit zusätzlichem Fokus auf der Silbenmethode unterrichtet wurde.

Die vorliegende Studie kann dabei als Interventionsstudie angesehen werden, da Tests mit Messwiederholungen – Einstufung vor Kursbeginn, Follow-up 1 nach 100 Kurslektionen, Follow-up 2 nach 200 Lektionen und die telc-A1-Prüfung nach 300 Kurslektionen – in allen Kursgruppen stattfanden (vgl. Marx 2012, S. 64; Albert/Marx 2016, S. 89). Durch dieses Vorgehen besteht zudem «eine grössere Sicherheit, dass die individuelle sprachliche Zuwendung im Rahmen eines Förderprogramms auf die Förderung anstatt auf andere Faktoren zurückzuführen ist» (Marx 2012, S. 66). Um die Standardisierung der Interventionsstudie zu gewährleisten (vgl. Darsow/Felbrich 2014, S. 236), musste die Treatmentgruppe 2 mindestens ein Viertel (d. h. über 10 Minuten pro Lektion) der Kurszeit mit der Silbenmethode verbringen. Dazu haben die jeweiligen Lehrpersonen täglich ein Silbenprotokoll (s. Anhang, Kap. 1, S. 213) ausgefüllt, das am Kursende der Forscherin abgegeben wurde. In diesem Protokoll wurde stichwortartig notiert, was genau gemacht wurde, wie lange die Übungen dauerten und welches Partnerformat verwendet wurde. Damit sollte nachgewiesen werden können, dass das Treatment planungskonform und in ausreichender Qualität von den Lehrpersonen durchgeführt worden ist.

3.2 Zweitschriftlernerurse

Die dieser Dissertation zugrundeliegenden Zweitschriftlernerurse wurden allesamt von der Erziehungsdirektion des Kantons Bern subventioniert. Ohne diese Unterstützung hätten die Kurspreise ungefähr das Doppelte betragen, was weder für die Kursteilnehmenden noch für die zuweisenden Stellen attraktiv gewesen wäre. Dank dieser Finanzierung konnte eine

vergleichsweise hohe Anzahl an Teilnehmenden für die Studie gewonnen werden. Allerdings mussten die Kurse von Institutionen angeboten werden, die den Subventionskriterien der Erziehungsdirektion des Kantons genügen. Mehrere Institutionen kamen infrage und eine ganze Reihe an Einrichtungen bekundeten Interesse am Forschungsprojekt. Letztlich wurden zwei Institutionen ausgewählt, die beide über einen grossen Erfahrungsschatz sowohl im Unterrichten von Alphabetisierungskursen sowie im A-Niveaubereich verfügen. Sie bleiben in der vorliegenden Untersuchung anonym und werden im weiteren Institution P und Institution Q genannt. Um den organisatorischen Aufwand für die Institutionen so gering wie möglich zu halten, wurden die Zweitschriftlernerkurse in das bestehende Kursprogramm integriert. Das bedeutet, dass Institution P etwas früher begonnen hat als Institution Q und zudem eine Woche Ferien hatte.

Bei beiden Institutionen wurden insgesamt drei Zweitschriftlernerkurse à zwölf Personen angeboten, was einer Gesamtteilnehmerzahl von 72 Personen entspricht. Dabei weicht die Höchstteilnehmerzahl von zwölf vom *Zweitschriftlernerkurs* in Deutschland ab, für den eine deutlich höhere Teilnehmerzahl von bis zu 25 festgelegt ist (s. Abschn. 2.1.1.2 und 8.3). Das Projekt orientiert sich hinsichtlich der Gruppengrösse am Rahmencurriculum (vgl. Lenz et al. 2009, S. 61) und an den Gegebenheiten der Schweiz.

Zur Einarbeitung der Lehrpersonen in die Silbenmethode wurde eine mündliche Einführung durch die Forschende gegeben und es wurden Unterrichtsbeispiele (s. Anhang, Kap. 2, S. 214) und der Artikel zu den drei Unterrichtsstufen der Silbenmethode von Rokitzki et al. (2013) zur Verfügung gestellt (s. Abschn. 1.4.1). Zusätzlich konnten die Lehrpersonen selbst Übungen entwerfen oder auf die Übungsauswahl aus Lehrwerken zurückgreifen. Damit die Lehrpersonen den Unterricht an ihr Kursniveau, an die Kursgruppe und an die individuellen Bedürfnisse der Teilnehmenden anpassen konnten, wurde auf eine starre Vorgabe der zu lösenden Übungen verzichtet. Festgelegt wurde jedoch, dass täglich mindestens ein Viertel der Kurszeit mit der Silbenmethode gearbeitet wird. Während des Kursverlaufs fanden zwei Austauschsitzungen mit den Lehrpersonen und der Forscherin statt, um die Möglichkeit zu geben, offene Fragen und Schwierigkeiten im Unterricht zu diskutieren.

Das Konzept für *Zweitschriftlernerkurse* (BAMF 2018a) beinhaltet einen wesentlichen Unterschied zu den regulären Alphabetisierungskursen der Schweiz, den die Lehrpersonen unabhängig vom Treatment zu beachten hatten: Lediglich die ersten 100 Kurslektionen gelten der Einführung in das deutsche Schriftsystem. Anschliessend wird mit dem Niveau A1

begonnen, wobei das langsamere Lese- und Schreibtempo der Lernenden gegenüber regulären A1-Lernenden berücksichtigt werden muss (vgl. BAMF 2018a, S. 19). In den Kursen wurde zudem verstärkt das Konzept der Grundbildung und Literalität miteinbezogen (s. Abschn. 2.1.3), anstelle des ansonsten üblichen Fokus auf die Sprache allein. Die Kursleitenden wurden beispielsweise ermutigt, gezielt Lernstrategien der Kursteilnehmenden zu fördern. Um die Kursinhalte nachvollziehbar zu gestalten, mussten alle Lehrpersonen der Forscherin am Kursende eine detaillierte Semesterplanung aushändigen.

Damit die Störvariablen (s. Abschn. 5.3.4) so gering wie möglich gehalten werden konnten, unterrichteten alle Kursgruppen mit dem gleichen Lehrwerk. Die Lehrpersonen konnten das Lehrwerk für die zweite und dritte Kursphase anlässlich der ersten Austauschsitzung gemeinsam auswählen, wobei die Wahl auf *Schritte plus Neu* (Niebisch et al. 2017) fiel. Mit diesem Lehrwerk hatten die Lehrpersonen bereits Erfahrung, da es in beiden Institution in regulären A1-Kursgruppen verwendet wird, was die Kursvorbereitung vereinfachte. Für die erste Kursphase arbeiteten zudem alle Kursgruppen mit *Schritte plus Alpha kompakt* (Böttinger 2017). Dieses Lehrwerk wurde von der Forscherin gewählt, da es kompatibel mit den Schritte-Büchern ist, die Teilnehmenden damit schon die Aufgabenstellungen der Bücher kennenlernen konnten und somit ein einfacherer Einstieg ins A1 stattfinden sollte.

Ziel war es, den Stoff des A1-Niveaus bis Kursabschluss in der Kursgruppe durchzunehmen, wobei es den Lehrpersonen freigestellt war, wie detailliert sie das Kursbuch dazu benutzten. So hat Institution Q in allen Kursgruppen den Teilnehmenden beide Lehrwerke (A1.1 und A1.2) abgegeben, wohingegen in Institution P nur das erste Buch (A1.1) ausgehändigt wurde.

3.3 Stichprobenbildung und Akteure

Eine Vollerhebung, mit der die Grundgesamtheit der Zweitschriftlernenden in der deutschsprachigen Schweiz hätte untersucht werden können, kam für das vorliegende Projekt nicht infrage. Erstens war vor Beginn der Studie ungewiss, ob das *Zweitschriftlernerkonzept* überhaupt ein gewinnbringendes Kursgefäß ist, weshalb eine Pilotstudie in einem einzelnen Kanton vorgezogen wurde. Zweitens war eine Vollerhebung im Rahmen einer Dissertation aus Ressourcengründen nicht möglich. Die Tests wurden zu allen Messzeitpunkten von der Forscherin durchgeführt, unter Einbeziehung von freiwilligen Helfern oder Studenten. Bei einer Ausdehnung auf weitere Kantone hätte die Forscherin sich nicht auf eine ähnliche Form der Freiwilligenarbeit stützen können.

Anstelle einer Vollerhebung wurde auf eine Stichprobe im Kanton Bern zurückgegriffen, die für die Grundgesamtheit der Zweitschriftlernenden in der deutschsprachigen Schweiz repräsentativ sein soll. Da die Grundgesamtheit aller Zweitschriftlernenden im Kanton Bern, wie auch in der Schweiz, unbekannt ist, wurde ein nicht zufallsbedingtes Stichprobenverfahren, präziser eine Quotenstichprobe mit dem Merkmal Zweitschriftlernende mit Zielsprache Deutsch, genutzt. Das heisst, die Teilnehmenden wurden nicht zufällig für die Studie ausgewählt, sondern es wurden gezielt Zweitschriftlernende für die Studie herangezogen. Ein solches Verfahren kann kein exaktes Miniaturbild der Grundgesamtheit in der deutschsprachigen Schweiz darstellen, dennoch sind vorsichtige Verallgemeinerungen möglich (vgl. Grum/Legutke 2016, S. 81–83).

Um das Merkmal Zweitschriftlernende für die Quotenstichprobe zu erfüllen, mussten die angemeldeten Teilnehmenden vor dem Einstufungstest einen Fragebogen in der Erstschriftsprache ausfüllen, in dem die Lese- und Schreibkompetenzen der Teilnehmenden überprüft wurden. Dabei mussten sie mindestens als funktional alphabetisiert eingestuft werden, d. h. in der L1 Niveau A2 oder höher aufweisen (vgl. Feldmeier 2015, S. 12–13). Die Bewertung erfolgte mit Hilfe von Übersetzern mit gleicher Erstsprache und aus demselben Herkunftsland (s. Abschn. 4.3.2). Dieses Vorgehen wurde dem blossen Abfragen der Schulbiographie im Herkunftsland vorgezogen, da Schule allein noch nicht als Garant für den Entscheid gelten kann, ob eine Person ausreichend hohe Lese- und Schreibkompetenzen aufweist (vgl. Feldmeier 2010c, S. 22). Ein zweites zu erfüllendes Merkmal um als Zweitschriftlerner zu gelten, war, dass die Person im lateinischen Schriftsystem Alphabetisierungsbedarf aufwies oder es nicht routinemässig beherrschte (vgl. BAMF 2018a, S. 12).

Sogenannte Gatekeepers oder auch Schlüsselpersonen bzw. Institutionen wie die Flüchtlingssozialdienste (Caritas und Schweizerisches Rotes Kreuz), das Programm Deutsch intensiv – Français intensif von HEKS, Institutionen, die Asylsuchende und Flüchtlinge betreuen (Durchgangszentren, Kompetenzzentrum Integration etc.), Sozialdienste für Flüchtlinge, die seit mehr als sieben Jahren in der Schweiz sind und diverse Institutionen, die Deutschkurse anbieten, haben in Zusammenarbeit mit der Autorin einen wesentlichen Beitrag beim Finden von Teilnehmenden geleistet, denn sie sind es, die die Kursteilnehmenden in den Deutschkursen anmelden.

3.3.1 Kurseinteilung

Um allfällige aus den Institutionen resultierende Störvariablen (s. Abschn. 5.3.4) zu neutralisieren, wurden die Treatments 1 und 2 in beiden Institutionen durchgeführt (vgl. Darsow/Felbrich 2014, S. 233). Die Kursverteilung pro Institution wird aus folgender Darstellung ersichtlich:

Institution P	Institution Q
Kurslevel hoch (P1), Treatment 2	Kurslevel hoch (Q1), Treatment 1
Kurslevel mittel (P2), Treatment 1	Kurslevel mittel (Q2), Treatment 2
Kurslevel tief (P3), Treatment 1	Kurslevel tief (Q3), Treatment 2

Tab. 3: Zweitschriftlernerkurse
Treatment 1: Zweitschriftlernerkonzept *ohne* Silbenmethode
Treatment 2: Zweitschriftlernerkonzept *mit* Silbenmethode

In Institution P wurde eine Kursgruppe mit und zwei ohne zusätzlichem Fokus auf der Silbenmethode unterrichtet (s. Tab. 3). In Institution Q wurde umgekehrt in zwei Kursen die Silbenmethode verwendet und in einem nicht. Die Kursgruppen wiesen bei beiden Institutionen drei verschiedene Levels auf: ein tiefes, ein mittleres und ein hohes.

Die Einteilung der 67 Lernenden erfolgte anhand des Einstufungstests. Dazu wurden die Lernenden anhand ihres Testergebnisses absteigend sortiert aufgelistet und in drei gleichgrosse Kurslevel eingeteilt. Dem hohen und dem mittleren Niveau wurden 22 Personen zugeordnet und der tiefen Gruppe 23. Dem tiefen Level wurde eine Person mehr zugeteilt, weil bei einem der Prüfungsteilnehmenden der Kursabbruch zu antizipieren war, da derjenige sich gerade in den Abschlussverhandlungen eines Arbeitsvertrags befand.

Innerhalb dieser drei gebildeten Kurslevel wurden die Teilnehmenden dann absteigend nach Prüfungsergebnis am Einstufungstest alternierend den beiden Institutionen zugeordnet.

Bei der Einstufung der Teilnehmenden in die Kursgruppen mussten jeweils zusätzlich Faktoren wie Kinderbetreuung und mögliche Kurszeiten berücksichtigt werden. Da lediglich eine Institution einen Kinderbetreuungsdienst anbot, mussten Teilnehmende mit Kindern speziell in dieser Institution angemeldet werden, was bei insgesamt neun Teilnehmenden zutraf. Auch die Kurszeiten variierten je nach Institution und Kurslevel (s. Tab. 4):

Kurszeit	Kurslevel	Kurszeit	Kurslevel
08:15–10:45	P1	09:00–11:35	Q1
08:15–10:45	P3	09:00–11:35	Q3
11:15–13:45	P2	14:00–16:35	Q2

Tab. 4: Kurszeiten Institution P und Q

Am Schluss wurde geprüft, ob in jedem Kurslevel die jeweiligen Kursgruppen bei den Institutionen einen ähnlich hohen Mittelwert beim Einstufungstest aufwiesen. Wie aus Tab. 5

ersichtlich, ist dies der Fall. Innerhalb der Kurslevel sind die Werte dicht beieinander oder im Falle des mittleren Niveaus gar identisch.

Kursgruppe	Mittelwert nach Einstufungstest
P1 (Kurslevel hoch)	57.4
Q1 (Kurslevel hoch)	55.4
P2 (Kurslevel mittel)	41.0
Q2 (Kurslevel mittel)	41.0
P3 (Kurslevel tief)	25.6
Q3 (Kurslevel tief)	25.7

Tab. 5: Mittelwerte der Zweitschriftlernerkerse
Punkteskala: 0 – 100

In den ersten Kurswochen sind insgesamt drei Personen nie im Kurs erschienen und zwei Personen wurden wieder abgemeldet. Danach war die endgültige Kursgruppengrösse die folgende: Auf dem oberen Niveau 10 Lernende bei Institution P und 11 bei Institution Q, auf dem mittlerem Niveau jeweils 10 Teilnehmende und auf dem tieferen Niveau 10 Personen bei Institution P und 11 bei Q (s. Abschn. 3.3.2).

3.3.2 Sample

An den Kursen teilnehmen konnten Zweitschriftlernende aus dem Kanton Bern mit der Zielsprache Deutsch. Als Kriterium galt, dass die Person in einem nichtlateinischen Schriftsystem schreiben und lesen kann und noch (deutliche) Schwierigkeiten im Umgang mit dem lateinischen Schriftsystem aufweist. Aufgrund der drei Kursniveaus konnten Zweitschriftlernende mit und ohne Vorkenntnisse(n) im lateinischen Schriftsystem am Kurs teilnehmen. Da das Kursziel mit dem aktuellen KIP übereinstimmt (s. Abschn. 2.2.1), war der Zustrom von Anmeldungen gross. Insgesamt wurden 80 Personen für die Kurse angemeldet. Nach der Einstufung im Juni 2018 wurden 13 Teilnehmer jedoch nicht für die Kurse zugelassen, da sie entweder in keiner Sprache (primäre Analphabeten) oder im lateinischen Schriftsystem lesen und schreiben konnten und somit nicht als Zweitschriftlernende gelten (s. auch Abschn. 4.3.2). Alle zugelassenen Personen und Institutionen wurden über die Studie informiert und die Teilnehmenden mussten ihr schriftliches Einverständnis für die Verwendung ihrer anonymisierten Daten für wissenschaftliche Zwecke geben. Um sicherzustellen, dass die Kursteilnehmenden die Angaben verstehen können, wurde dieser Teil des Fragebogens in die jeweilige Erstschriftsprache übersetzt.

Insgesamt konnten 67 Teilnehmende auf die sechs Kurse verteilt werden. Davon sind jedoch drei Teilnehmende nie im Kurs erschienen und zwei wurden wieder abgemeldet. Von den restlichen 62 Teilnehmenden wurden schlussendlich 57 Lernende in die Datenauswertung

aufgenommen, da nur Personen berücksichtigt wurden, die bei allen Erhebungen inkl. der telc-Prüfung anwesend waren. Drei der fünf nicht berücksichtigten Fälle haben aus persönlichen Gründen den Kurs abgebrochen und zwei sind nicht zum telc-Test erschienen. Die im weiteren diskutierten Angaben stammen aus der Auswertung der Lernerfragebögen (s. Abschn. 4.3.2).

Wie aus Abb. 1 ersichtlich, stellen Eritreer mit 39 Teilnehmenden (68 % der Stichprobe) die am häufigsten vertretene **Nationalität** dar, gefolgt von elf Personen aus Syrien (19 %). Aus Afghanistan sind drei Lernende (5 %), aus Sri Lanka und Tibet (Volksrepublik China) jeweils zwei Teilnehmende (je 4 %) in dieser Studie vertreten.

Analog zur Nationalität differiert die Verteilung der **Erstsprachen**. Der Grossteil der Probanden, nämlich 37 Personen (65 %), hat Tigrinya als Erstsprache, gefolgt von Arabisch mit sieben Teilnehmenden (12 %), Kurdisch mit fünf (9 %) und Persisch mit drei Teilnehmenden (5 %). Die Gruppe mit Tibetisch und Tamil als Erstsprache ist jeweils gleich gross und mit je zwei Teilnehmenden (4 %) vertreten. Eine Person (1 %) hat zudem Saho als Erstsprache.²⁵

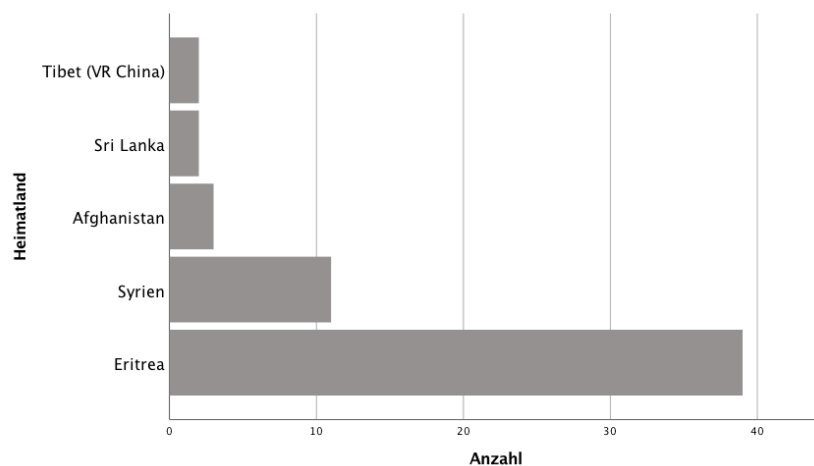


Abb. 1: Heimatländer der Teilnehmenden

Zwischen dem Heimatland und der Erstsprache besteht ein deutlicher positiver Zusammenhang. Die beiden Variablen korrelieren hochsignifikant mit einem Koeffizienten von $r = .84$ ($p = .000$), was naheliegend ist, da die Erstsprache häufig vom Heimatland abhängt. Wie in Abschnitt 1.1.2 besprochen, entspricht zudem die Erstsprache nicht immer der Erstschriftsprache. So nennt ein eritreischer Teilnehmer Saho als Erstsprache, ist jedoch in

²⁵ Die Prozentsumme beträgt nicht genau 100 %, was sich auf die Rundung der Dezimalwerte zurückführen lässt. Gleiches gilt für den übrigen Text.

Arabisch alphabetisiert. Das gilt auch für die fünf Personen mit Kurdisch als Erstsprache, die ebenfalls in Arabisch alphabetisiert sind.

Die **Geschlechterverteilung** ist nicht ausgeglichen, so befinden sich unter den Teilnehmenden mehr Männer ($n = 35$) als Frauen ($n = 22$). In Abb. 2 ist die Verteilung des Geschlechts nach Geburtsjahr dargestellt. Der Grossteil der Teilnehmenden ist zwischen 20 bis 29 Jahre alt (49 %, $n = 28$). Die zweitgrösste Gruppe bilden die 30- bis 39-Jährigen mit 16 Teilnehmenden (28 %) und die drittgrösste Gruppe die 40- bis 49-Jährigen mit zehn Lernenden (18 %). Dazu kommen eine 19-jährige Teilnehmerin und zwei 51-jährige Teilnehmer.

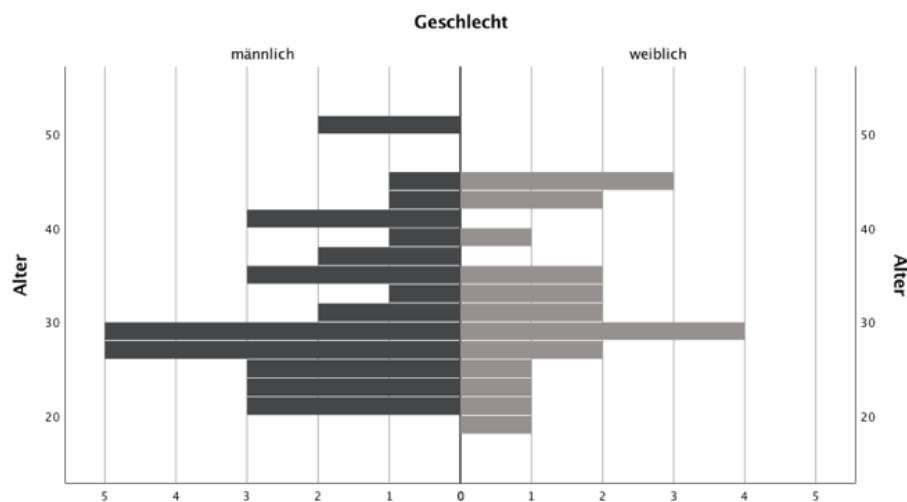


Abb. 2: Geschlechterverteilung

Die Mehrheit der Teilnehmenden hat im jeweiligen Heimatland sieben bis neun **Schuljahre** ($n = 29$) absolviert (s. Abb. 3). Die Anzahl Schuljahre der übrigen Teilnehmenden ist relativ gleichmässig verteilt: 10 bis 12 Schuljahre ($n = 9$), 4 bis 6 Schuljahre ($n = 13$) und 1 bis 3 Schuljahre ($n = 5$). Lediglich eine Person gab an, keine Schule im Heimatland besucht zu haben. Nur eine Person hat einen tertiären Abschluss einer **Universität** und neun weitere Teilnehmende haben eine **Berufslehre** oder eine ähnliche Ausbildung abgeschlossen. Mehr als 60 % aller Personen ($n = 35$) haben angegeben, in ihrem Heimatland einer Arbeit nachgegangen zu sein.

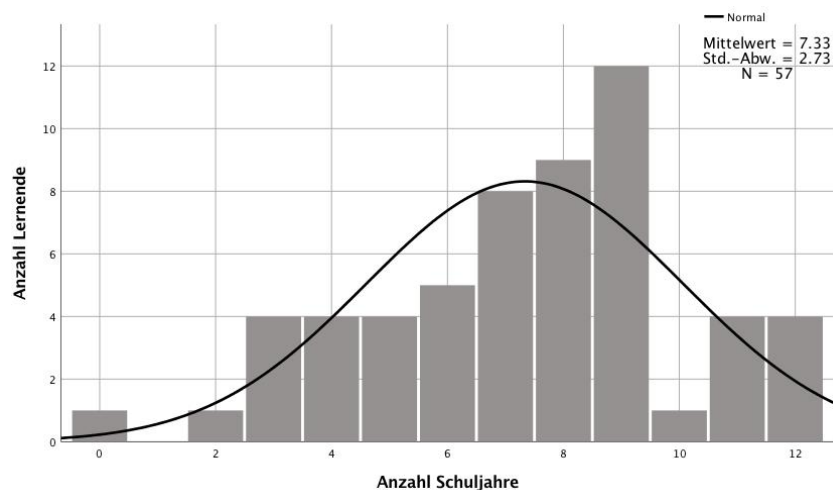


Abb. 3: Anzahl der besuchten Schuljahre im Heimatland

Von den insgesamt 57 Teilnehmenden haben 46 Personen vor Projektteilnahme bereits mindestens einen **Deutschkurs** besucht. Davon haben 18 Teilnehmende einen regulären Deutschkurs ohne speziellem Fokus auf Lesen und Schreiben (z. B. A-Niveau) und 28 Teilnehmende einen Alphabetisierungskurs absolviert.

Insgesamt 13 Teilnehmende gaben an, mindestens eine **Fremdsprache** zu beherrschen, davon fünf Personen Englisch, vier Arabisch und jeweils eine Person Amharisch oder Chinesisch. Zwei Personen nannten zwei Fremdsprachen: Amharisch und Arabisch bzw. Amharisch und Englisch. Alle Personen, die Englisch als Fremdsprache angaben, weisen nur mündliche Kenntnisse auf, weshalb die Definition als Zweitschriftlerner nach wie vor zulässig ist. Die Teilnehmenden mit Fremdsprachenkenntnissen stammen aus Eritrea, Syrien und Tibet.

3.3.3 Lehrpersonen

Die Lehrpersonen sind alle durch die pädagogischen Leitungen der Institutionen über die Pilotkurse informiert worden und konnten sich freiwillig für das Unterrichten in den Zweitschriftlernerkursen melden. Dabei fanden sich schnell Lehrpersonen, die Interesse am Projekt bekundeten und einen Kurs übernehmen wollten. Bei Institution Q im Kurs Q3 wurde eine Lehrperson speziell für das Projekt neu angestellt. Die Auswahl der Lehrpersonen erfolgte durch die pädagogische Leitung der Institutionen, wobei es bei beiden Institutionen während des Kursverlaufs einen Lehrpersonenwechsel gab. So musste bei Institution Q eine Lehrperson aus persönlichen Gründen den Kurs nach drei Wochen abgeben. Im Kurs P2 bei Institution P, der von zwei Personen alternierend unterrichtet wurde, kündigte eine der beiden Lehrpersonen nach den Herbstferien, woraufhin sich auch ihre Kollegin vom Kurs zurückzog und neue Lehrpersonen eingesetzt wurden.

Die vor Kursbeginn geäusserten Einschätzungen zum Kurskonzept waren mit einer Ausnahme positiv: Nur eine Lehrperson war der Meinung, dass es nicht möglich sei, innerhalb von 300 Kurslektionen das Zielniveau A1 zu erreichen.

In Tab. 6 und Tab. 7 ist die detaillierte Übersicht der Kursleitenden ersichtlich, damit der Faktor «Qualifizierung der Lehrpersonen» in der Diskussion berücksichtigt werden kann. Damit die Anonymisierung gewahrt bleibt, wurden nichtdeutsche Erstsprachen als *Andere* kategorisiert und die Ausbildung der Kursleitenden weggelassen, damit keine Rückschlüsse auf die Personen möglich sind. Die Zuordnung in die Zusatzqualifizierungsmassnahmen gemäss BAMF (vgl. BAMF 2018d) geschah unter bestmöglicher Berücksichtigung der verfügbaren Informationen.

Kurs	Erstsprache	Kurserfahrung in Jahre	Benötigter Qualifizierungslehrgang nach BAMF
P1	Andere	DaZ: 10–15 Alpha: 0	keine Zusatzqualifizierung
P2; im Oktober aufgehört	Deutsch/Andere	DaZ: 2–3 Alpha: 1	unverkürzte Zusatzqualifizierung
	Deutsch	DaZ: 5–10 Alpha: 1	keine Zusatzqualifizierung
P2; im Oktober angefangen	Deutsch	DaZ: 4–5 Alpha: 1	unverkürzte Zusatzqualifizierung
	Deutsch	DaZ: > 15 Alpha: 5–10	unverkürzte Zusatzqualifizierung
P3	Andere	DaZ: > 15 Alpha: 2–3	keine Zusatzqualifizierung

Tab. 6: Lehrpersonen Institution P

Kurs	Erstsprache	Kurserfahrung in Jahre	Benötigter Qualifizierungslehrgang nach BAMF
Q1; im August aufgehört	Deutsch	DaZ: > 15 Alpha: 4–5	unverkürzte Zusatzqualifizierung
Q1; im August angefangen	Andere	DaZ: 2–3 Alpha: < 1	keine Zusatzqualifizierung
	Andere	DaZ: 5–10 Alpha: 4–5	keine Zusatzqualifizierung
Q2	Andere	DaZ: 5–10 Alpha: 1	keine Zusatzqualifizierung
	Andere	DaZ: 10–15 Alpha: 1	keine Zusatzqualifizierung
Q3	Deutsch	DaZ: 10–15 Alpha: 10–15	keine Zusatzqualifizierung

Tab. 7: Lehrpersonen Institution Q

3.4 Forschungsethik beim Testen von Flüchtlingen

Empirische Forschung mit Flüchtlingen kann aus verschiedenen Gründen eine Herausforderung darstellen (vgl. Röder et al. 2018, S. 314). Besonders betrifft das Asylsuchende mit Status N. So ist beispielsweise eine Langzeitstudie mit dieser Zielgruppe schwierig, da sich der Aufenthaltsstatus jederzeit ändern kann und somit die Zugehörigkeit

zur definierten Population. Auch ist mit einer erhöhten Drop-out-Quote infolge der Rückkehr ins Heimatland bei negativem Asylentscheid oder mit Wegzug aus dem Einzugsgebiet eines Sprachkurses bei positivem Asylentscheid zu rechnen. Solange der Aufenthaltsstatus von Asylsuchenden ungewiss ist, müssen zudem die gegenüber den Forschenden geäusserten Aussagen unter Vorbehalt betrachtet werden, da davon auszugehen ist, dass Personen mit Status N bewusst Antworten geben, von denen sie denken, dass sie ihren Aufenthaltsstatus nicht beeinträchtigen oder diesen gar begünstigen können (vgl. Röder et al. 2018, S. 318–319). Aus diesen Gründen werden Asylsuchende mit Ausweis N nicht in die vorliegende Studie miteingeschlossen.

Als grosse Herausforderung stellte sich in dieser Studie der Fragebogen heraus, der in fünf Sprachen übersetzt werden musste. Dabei waren nicht nur die Erstschriftsprachen aller Kursteilnehmenden zu berücksichtigen, sondern es musste auch auf unterschiedliche Bildungshintergründe Rücksicht genommen und daher eine einfache Sprache verwendet werden. Zudem sollten die Fragen in allen Sprachen dieselbe Bedeutung haben und keine allzu deutlichen inhaltlichen Unterschiede aufweisen. Dies konnte erreicht werden, indem die Übersetzung der Fragebögen selbstständig durch erfahrene Übersetzer erfolgte und darüber hinaus gemeinsam mit der Forscherin hinsichtlich semantischer Feinheiten hinterfragt wurde. Zusätzlich wurde jeder Fragebogen durch einen weiteren Übersetzer überprüft, um mögliche Verständnisschwierigkeiten zu identifizieren. Diese Überarbeitungen ergaben schlussendlich ausreichend einfache Fragesätze, die von allen Kursteilnehmenden verstanden werden konnten (s. Abschn. 4.3.2).

Ein weiterer Punkt, der in der Forschung mit Flüchtlingen beachtet werden muss, ist der umsichtige Umgang mit Antwortskalen bei quantitativen Erhebungen. Die Mehrheit der Flüchtlinge hat in ihrem bisherigen Leben wenig Erfahrungen mit dem Ausfüllen von Fragebögen gemacht, was dazu führt, dass sie die Antwortskalen nicht kennen und nicht wissen, was damit verlangt wird. So gab beispielsweise der Übersetzer für Arabisch an, dass er in Syrien mit fünfstufigen Antwortskalen eher schlechte Erfahrung gemacht habe, da dies eine zu grosse Auswahl für die Befragten darstelle. Auf dieses Problem weisen auch Röder et al. (2018) hin. Im Fragebogen wurden dennoch fünfstufige Antwortskalen verwendet, da die restlichen Übersetzer davon abgeraten hatten, eine geringere Anzahl an Antwortmöglichkeiten zu benutzen. So wird gemäss ihren Überlegungen bzw. Erfahrungen bei

zu wenig ankreuzbaren Feldern die Idee, dass es sich um ein Spektrum von Antwortmöglichkeiten handelt, nicht ausreichend ersichtlich.

Flüchtlinge zeigen des Weiteren ein teilweise deutlich ausgeprägtes Misstrauen gegenüber Interviewenden, was sich wiederum auf die Beantwortung der Fragen auswirken kann (vgl. Röder et al. 2018, S. 322–323). Aus diesem Grund ist es von besonderer Bedeutung, mitzuteilen, dass die Teilnahme an der Studie freiwillig ist und die Daten streng vertraulich behandelt werden. Zur Ermittlung der Schreibkompetenzen der Teilnehmenden in der Erstschriftsprache sollten diese am Schluss des Lernerfragebogens einen kurzen Text über sich oder ihre Familie schreiben. Auch wenn klar kommuniziert wurde, dass der Inhalt fiktiv sein darf, hat sich die Wahl des Themas *Familie* nachträglich als nicht optimal herausgestellt. Ein Teilnehmer weigerte sich anfangs strikt, sich zu diesem Punkt zu äussern und kam der Bitte erst nach entsprechender Erläuterung und Zuspruch durch einen Mitteilnehmenden nach. Daher ist für andere Forschungen die Wahl eines neutraleren Themas zu empfehlen.

Ein weiterer, grundsätzlich relevanter Punkt in der Forschung mit Flüchtlingen (und mit Menschen überhaupt) besteht darin, die teilnehmenden Personen nicht nur als Testsubjekte oder gar -objekte zu verstehen. Auch wenn die Studien und Testverfahren möglichst hohen akademischen Standards genügen müssen, dürfen darunter der zwischenmenschliche Aspekt und der Respekt vor dem Individuum nicht leiden. Gerade Flüchtlinge stellen eine vulnerable Gruppe dar, da die Gründe für ihre Flucht und diese selbst häufig traumatisch sind. Umso wichtiger ist es, von Anfang an eine Vertrauensbasis herzustellen: Das kann durch die Bereitstellung männlicher *und* weiblicher Prüfender geschehen, damit weibliche Teilnehmende, wenn dies gewünscht ist, von weiblichen Prüfenden getestet bzw. befragt werden können, oder durch erstsprachliche Begleiter, die Hilfestellungen geben können. Hier ist jedoch zu beachten, dass Letzteres auch kontraproduktiv sein kann, da Flüchtlinge vielfach aus Ländern kommen, in denen die Zivilgesellschaft gespalten ist, sodass die Begleitperson möglicherweise andere politische Ansichten vertritt als die zu testende Person, was sich dann wiederum negativ auf deren Offenheit auswirken kann. Es ist zudem von Vorteil, wenn die Forschenden die kulturellen Gegebenheiten der Teilnehmenden kennen und hohe interkulturelle Kompetenzen aufweisen, damit sich die Teilnehmenden in ihrer Kultur und Tradition wertgeschätzt fühlen; so zeugt es beispielsweise eritreischen Teilnehmenden von Respekt, wenn beim Händeschütteln der linke Arm den rechten Unterarm stützt. Auch erweist es sich als hilfreich, wenn die Forschenden einige Wörter und Phrasen in der Erstsprache der

Teilnehmenden beherrschen, damit sie bei Fragestellungen helfen können und weil dies bei den Teilnehmenden die Bereitschaft zu einer unvoreingenommenen Forschungsteilnahme erhöhen kann.

3.5 Forschungsfragen

Wie bereits in Abschnitt 1.3 beschrieben, ist im deutschsprachigen Raum ein Trend zu einer deutlicheren Differenzierung der Teilnehmer in Alphabetisierungskursen zu beobachten. So wurde zum Beispiel im *Lernraum* in Wien im Jahr 2016 ein Pilotprojekt für Zweitschriftlernende durchgeführt. In Deutschland werden seit Februar 2017 *Zweitschriftlerner*kurse (BAMF 2018a) vom BAMF gefördert; einerseits finanziell und andererseits anhand eines spezifizierten Konzepts für Zweitschriftlernende. Diese Trennung der beiden Gruppen, Zweitschriftlernende und primäre Analphabeten, scheint aus mehreren Gründen – wie in Abschnitt 1.3 ausgeführt – sinnvoll. Dennoch wird in der Schweiz der Tatsache, dass Zweitschriftlernende eine schnellere Progression durchlaufen können als primäre Analphabeten, bisher nicht Rechnung getragen. Auch bestehen Forschungslücken im Bereich der Zweitspracherwerbsforschung mit Flüchtlingen, insbesondere bezüglich des Schriftspracherwerbs.

Ziel dieser Forschungsarbeit ist die Untersuchung von Zweitschriftlernenden in einem für sie speziell konzipierten Kurssystem. Daraus ergeben sich zwei Forschungsfragen:

- F1. Erreichen Zweitschriftlernende, die in der deutschsprachigen Schweiz nach dem Deutschen Zweitschriftlernerkonzept unterrichtet werden, innerhalb von 300 Kurslektionen ein A1-Niveau nach telc?
- F2. Welche Unterschiede sind in verschiedenen Teilbereichen der Sprachkompetenz feststellbar, wenn Zweitschriftlernende im Rahmen des Zweitschriftlernerkonzepts aus Deutschland mit oder ohne speziellem Fokus auf der Silbenmethode unterrichtet werden?

3.6 Hypothesen

Das Ziel dieser Forschungsarbeit ist die Prüfung der beiden Forschungsfragen und der daraus abgeleiteten Hypothesen.²⁶

- H1. Mindestens die Hälfte der Kursteilnehmenden der Treatmentgruppen 1 und 2 erreichen das Niveau A1 nach telc innerhalb von 300 Kurslektionen.
- H2. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im telc-Resultat.
 - H2.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.

²⁶ Zur Erinnerung: Treatmentgruppe 1 umfasst alle Kurse *ohne* Silbenmethode, Treatmentgruppe 2 die Kursgruppen *mit* zusätzlichem Fokus auf der Silbenmethode.

- H2.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.
- H2.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen.
- H2.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.
- H3. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im Lernerfolg²⁷.
 - H3.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.
 - H3.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.
 - H3.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen.
 - H3.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.
- H4. Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat und den Gesamtlernerfolg²⁸.
 - H4.1 Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat.
 - H4.2 Die Anzahl der Schuljahre hat einen Einfluss auf den Gesamtlernerfolg.
- H5. Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das Resultat im telc-Test und den Gesamtlernerfolg.
 - H5.1 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das telc-Resultat.
 - H5.2 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst den Gesamtlernerfolg.
- H6. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich des Arbeitsgedächtnisses.
- H7. Das Arbeitsgedächtnis wirkt sich auf alle vier Fertigkeiten aus.
 - H7.1 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Lesen.
 - H7.2 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Schreiben.
 - H7.3 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Sprechen.
 - H7.4 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Hören.

Die Hypothese H1 dient zur Überprüfung der ersten Forschungsfrage und H2 und H3 beantworten die zweite Forschungsfrage. Besonders relevante Definitionsmerkmale von Zweitschriftlernenden sind die Anzahl absolvierter Schuljahre im Heimatland und die schriftsprachlichen Kompetenzen in der Erstschriftsprache: H4 und H5 ermöglichen Aussagen darüber, wie gross der Einfluss dieser beiden Komponenten auf den Lernerfolg und das Abschneiden beim Abschlusstest A1 ist. Hypothese H6 überprüft den Einfluss der Silbenmethode auf das Arbeitsgedächtnis (gemessen mit dem Mottier-Test) und H7 soll zeigen, wie sich diese Unterschiede auf die vier Fertigkeiten auswirken.

²⁷ Der Lernerfolg entspricht der Lernprogression der Lernenden in den einzelnen Kursphasen (à 100 Kurslektionen). Lernerfolg 1 entspricht somit der Lernprogression von Lektion 1 bis Lektion 100, Lernerfolg 2 der Lernprogression zwischen Lektion 101 und Lektion 200 (s. Kap. 5.3).

²⁸ Der Gesamtlernerfolg gibt die Lernprogression der Lernenden zwischen Beginn der Studie und dem zweiten Messzeitpunkt nach 200 Kurslektionen wieder (s. Kap. 5.3).

Im Rahmen der Ausführungen zur Testkonstruktion im nächsten Kapitel soll auf die Operationalisierung der für die Hypothesenprüfung wichtigen Begrifflichkeiten eingegangen werden, bevor deren Kodierung in Kapitel 5 vorgestellt wird.

4. Testkonstruktion

Um festzustellen, wie sich die Kursteilnehmenden hinsichtlich ausgewählter Teilbereiche der Sprachkompetenz unterscheiden, wurde ein circa 50-minütiges Testverfahren entwickelt, das durch einen Lerner- und einen Lehrerfragebogen ergänzt wurde. Das Testverfahren besteht aus fünf Instrumenten: Elicited Imitation für die Fertigkeit Sprechen (s. Abschn. 4.2.1), Kunstwörter für das Arbeitsgedächtnis (s. Abschn. 4.2.2), Oral Reading Fluency für die Leseflüssigkeit (s. Abschn. 4.2.3) sowie ein Hörverstehenstest (s. Abschn. 4.2.4) und ein Diktat für die Laut-Buchstaben-Zuordnung und die Verständlichkeit (s. Abschn. 4.2.5). Was genau mit Hilfe der genannten Tests gemessen werden soll und wie die Lernerleistungen bewertet wurden, wird in diesem Kapitel dargestellt.

4.1 Kommunikative und allgemeine Sprachkompetenz

Um *fremdsprachliche Lernerleistungen* erheben, messen und beurteilen zu können, richten sich die in der vorliegenden Studie verwendeten Tests auf die Feststellung der Sprachkompetenz der Lernenden. Diese wird hier im Sinne von Bachman und Palmer (1990) als wesentliche Voraussetzung für den Gebrauch einer Sprache verstanden. Sprachkompetenz wird je nach Disziplin, beispielsweise in der Linguistik, der Psycholinguistik oder den Erziehungswissenschaften, unterschiedlich definiert, wobei sie stets eine Bandbreite an spezifischen Kompetenzen umfasst. In dieser Arbeit wird Sprachkompetenz vor dem Hintergrund des Sprachenlehrens und -lernens verortet (vgl. Jude/Klieme 2006, S. 9–10). Das heisst, es gilt nicht das Prinzip der vollständigen Erfassung der Sprachkompetenz, die allgemeine Fähigkeiten, motivationale Voraussetzungen, anforderungsbezogene Handlungsfelder oder andere Aspekte einschliesst (vgl. ebd., S. 11). Vielmehr sollen die gewählten Fertigkeiten – Sprechen, Hören, Schreiben, Lesen – die curricularen Schwerpunkte auch im Hinblick auf standardisierte Tests wie die telc-Prüfung A1 beleuchten. Aus dem Abschneiden der Teilnehmenden in diesen Teilkompetenzen soll ein adäquates Abbild ihrer Beherrschung des Deutschen ausserhalb der Testsituation geliefert werden (vgl. ebd., S. 9–10). Die Sprachkompetenz lässt sich als mehrdimensionales, aus verschiedenen empirisch überprüfbaren Teilkompetenzen bestehendes Konstrukt darstellen (vgl. ebd., S. 11). Die hier eingesetzten Tests sollen einerseits gewisse Teilaspekte sichtbar und vergleichbar machen und andererseits in ihrer Summe einen Eindruck über die allgemeine Sprachkompetenz (im

Unterschied etwa zu berufsspezifischen Sprachkompetenzen) der Teilnehmenden ermöglichen.

Die forschergenerierten Testinstrumente und die telc-Prüfung A1 messen beide die allgemeine Sprachkompetenz. Im Gegensatz zu den eigens für diese Studie entwickelten Tests erfasst die telc-Prüfung jedoch insbesondere die kommunikative Sprachkompetenz (vgl. Trim et al. 2009, S. 21). Seit Beginn der 70er Jahren, in denen die kommunikative Wende stattfand, wird die kommunikative Sprachkompetenz als übergeordnetes Lernziel festgelegt (vgl. Bachman 1990, S. 81–82). Die telc-Prüfung bezieht sich dabei auf den GER (vgl. telc gGmbH 2010, S. 5) und versteht darunter die Befähigung «zum Handeln mit Hilfe spezifisch sprachlicher Mittel» (Trim et al. 2009, S. 21).

Obwohl der telc-Test und die eigens für die Studie entwickelten Tests jeweils die allgemeinen Sprachkompetenzen messen, sind sie nicht direkt vergleichbar, da telc stärker die kommunikativen Kompetenzen fokussiert. Die Korrelation der gemessenen Fertigkeiten stellt sich dennoch, mit Ausnahme des Lesens, als mittel bis stark heraus (s. Abschn. 6.1.3.3).²⁹

4.2 Testinstrumente

Für die Forschungsarbeit konnte nicht auf bereits existierende Tests zurückgegriffen werden. Der Forscherin ist lediglich ein einziger Test bekannt, der speziell für Zweitschriftlernende entwickelt wurde: Das vom BAMF in Zusammenarbeit mit dem Goethe-Institut entwickelte *Einstufungssystem für die Integrationskurse in Deutschland* (Perlmann-Balme/Dengler 2007). Dieser Test konnte aber mangels Einwilligung des BAMF nicht benutzt werden. Aus diesem Grund wurden eigene Tests für die hier vorliegende Zielgruppe entwickelt. Diese dienen der Untersuchung der Treatments: des *Zweitschriftlernerkonzepts* und der Silbenmethode. Im Kontext der Fremdsprachenforschung und der vorliegenden Forschungsarbeit wird ein Test «zur Erhebung, Messung und Beurteilung fremdsprachlicher Lernerleistungen eingesetzt» (Harsch 2016, S. 205).

Um die sprachlichen Leistungsunterschiede von Zweitschriftlernenden zu erfassen, wurden vier der fünf eingesetzten Testinstrumente durch die Forscherin entwickelt, die jeweils den Fokus auf ein spezifisches Phänomen legen. Diese vier Tests betreffen die vier Fertigkeiten – Sprechen, Hören, Schreiben und Lesen. Dazu kommt ein von Grete Mottier (1951)

²⁹ Es ist anzunehmen, dass die forschergenerierten Tests, mit Ausnahme des Hörverstehens, eher auf sprachlichem Grundlagenwissen beruhen und einen trait-based-approach verfolgen, die Aufgaben der telc-Prüfung hingegen eher einem aufgabenbasierten Ansatz mit realem Lebenskontext, einem tasked-based-approach, nahekommen (vgl. National Academics of Sciences, Engineering and Medicine 2020, S. 22–25).

entwickelter Test zur Überprüfung des Arbeitsgedächtnis. Alle hier eingesetzten Tests dienen dem summativen Testen, das feststellen will, was die Lernenden zu einem bestimmten Zeitpunkt können (vgl. Harsch 2016, S. 206). Die Testevaluation besteht aus einer Mischung aus objektiver und subjektiver Bewertung (vgl. Grotjahn 2010, S. 213). Die Evaluation des Hörverstehens kann als objektiv bezeichnet werden, da die Auswertung mittels einer binären Unterscheidung von richtig/falsch geschieht und somit keine komplexe Evaluationsleistung vorliegt. Die übrigen entwickelten Sprachtests (Elicited Imitation, Oral Reading Fluency, Diktat) und die Kunstwörter nach Mottier zählen zu den Tests mit subjektiver Bewertung, da ihre Evaluation komplexer und somit subjektiver ist.

Um die Interrater-Reliabilität bei den vier Tests mit subjektiver Evaluationskomponente zu gewährleisten, wurde jede Bewertung zweifach, einmal durch die Forscherin und einmal durch eine weitere erfahrene Deutschkursleiterin, durchgeführt. Die Bewertungen fanden an gemeinsamen Sitzungen statt, um das Tempo und die Genauigkeit der Bewertungen zu kontrollieren. Falls die Bewertungen nicht identisch waren, wurde ein dritter Bewerter hinzugezogen. Alle Bewertenden hatten sich an das jeweilige Evaluationsraster zu halten und wurden vor der Bewertung geschult. In der Schulung wurde den Bewertenden zuerst das Evaluationsraster erklärt und allfällige Fragen beantwortet. Danach erhielten diese zu jedem der vier Tests Beispiele, die zusammen mit der Forscherin bewertet wurden. Die Bewertungen wurden anschliessend verglichen und diskutiert. Diese Beispiele dienten sodann als Standard. Die Bewertenden wurden zusätzlich kontrolliert indem die Retest-Items anhand ihrer Punktezahl verglichen wurden und das Beurteilungsverhalten pro Test auf Konsistenz überprüft wurde (in Anlehnung an Association of Language Testers in Europe (ALTE) 2006, S. 13–14).

Damit das Testverfahren primär Wissen und Können prüft, das die Lernenden in den Zweitschriftlernerkursen erworben haben können, wurde der Fokus auf das Niveau A1 gelegt. 30 % der Items stammen zudem aus dem A2-Niveau, um auch fortgeschrittene Lernende unterscheiden zu können. Alle Aufgabenformate wurden entweder in Studien mit erwachsenen Zweitschriftlernenden und/oder mit jugendlichen DaZ-/ESL-Lernenden eingesetzt, wodurch eine gewisse Gewähr besteht, dass sie mit der Zielgruppe durchführbar sind. Die für den Elicited Imitation und die Diktate benutzten Wörter entsprechen der

Häufigkeitsliste (A1- und A2-Niveau) von Tschirner (2016)³⁰ und die Sätze wurden, soweit möglich, in Anlehnung an den Alltag und die Lebenswelt der Teilnehmenden entwickelt. Das Hörverstehen und der Lesetext für den Oral Reading Fluency wurden mit freundlicher Genehmigung der jeweiligen Verlage aus A1-Lehrwerken kopiert. Die Kunstwörter entstammen dem Original Mottier-Test, der für Kinder und Jugendliche im DaZ-Bereich benutzt wird (vgl. Gamper et al. 2012, S. 5; Wild/Fleck 2013).

In Tab. 8 sind wesentliche Informationen über das Testverfahren schematisch zusammengefasst. Das Testverfahren wurde in zwei Teile gegliedert: Der erste, mündliche Teil wurde mit den Teilnehmenden einzeln durchgeführt und der zweite, schriftliche Teil gemeinsam mit der gesamten Kursgruppe. Auf die einzelnen Testinstrumente wird nun vertieft eingegangen.

Teil 1	Bsp. einer Forschung	Testfokus	Testkonstrukt / getesteter Bereich	Testmaterial	Bewertung
Elicited Imitation (Task)	Zweitschrift-lernende ESL (vgl. Ortega et al. 2002).	Unterschiedlich lange Äusserungen verstehen und sie in Form und Inhalt möglichst genau wiedergeben.	Mündliches Sprachverstehen und Sprachproduktion Niveau A1/A2.	4–16 Silben in insgesamt 30 Äusserungen.	Vollständigkeit und Genauigkeit der wiedergegebenen Äusserung.
Kunstwörter nach Mottier	Kinder/Jugendliche inkl. DaZ (vgl. Wild/Fleck 2013).	Gehörte Kunstwörter möglichst genau mündlich wiedergeben.	Arbeitsgedächtnis; phonologische Schleife.	30 Kunstwörter nach Mottier.	Genauigkeit der wiedergegebenen Wörter.
Oral Reading Fluency	Zweitschrift-lernende ESL (vgl. Lems 2003).	Genaueres, automatisiertes und sinnkonstituierendes lautes Lesen.	Lese流利igkeit Niveau A1.	Lehrwerkstext A1.	Genauigkeit der Lese流利igkeit.
Teil 2	Bsp. einer Forschung	Testfokus	Testkonstrukt / getesteter Bereich	Testmaterial	Bewertung
Hörverstehen	Kein Beispiel bekannt.	Verstehen der Hauptinformationen. Spezifische Informationen aus dem gehörten Text entnehmen. Details in einem gehörten Text verstehen.	Hörverstehen Niveau A1.	Lehrwerkstext A1, insgesamt 18/19 Aufgaben.	Korrektheit der verstandenen Informationen.
Diktat	Jugendliche ESL (vgl. Kazazoglu 2013).	Anwendung der Laut-Buchstaben-Zuordnung im Deutschen. Wörter und Sätze korrekt und verständlich verschriftlichen.	Laut-Buchstaben-Wissen Niveau A1/A2.	22 Wörter und 8 Sätze im A1/A2-Bereich.	Vollständigkeit der Laut-Buchstaben-Zuordnung und Verständlichkeit der wiedergegebenen Äusserung.

Tab. 8: Testverfahren

³⁰ Die Häufigkeitsliste beruht auf korpuslinguistischen Arbeiten des Herder-Instituts der Universität Leipzig (Jones/Tschirner 2006). Das Korpus ist ausgewogen und repräsentativ und es umfasst rund 4,2 Millionen Wörter. Das Verhältnis beträgt 70:20:10 für deutsche, österreichische und Schweizer Texte (vgl. Tschirner 2016, S. 4).

4.2.1 Elicited Imitation Task

Elicited Imitation Task oder kurz Elicited Imitation (EI), wird seit geraumer Zeit in der Spracherwerbsforschung als Testinstrument benutzt, vorwiegend in drei Hauptbereichen: in der Spracherwerbsforschung bei Kindern, in der Neuropsychologie und in der Zweitspracherwerbsforschung (vgl. Vinther 2002, S. 54). Da für diese Forschungsarbeit nur Letztere relevant ist, wird auf die Verwendung der Elicited Imitation in den anderen beiden Disziplinen nicht näher eingegangen.

Mit der Methode der Elicited Imitation wird der zu Prüfende aufgefordert, Stimuli in Form akustischer Äusserungen möglichst lautgetreu zu wiederholen (vgl. Winkler 2011, S. S. 528). Gemäss Vinther (2002, S. 62) beinhaltet Elicited Imitation verschiedene sprachliche Verarbeitungsebenen: Die Äusserung wird gehört, im Arbeitsgedächtnis dekodiert, verarbeitet, erinnert und in genau gleicher linguistischer Form wiedergegeben. Weniger klar ist, welche dieser kognitiven Aktivitäten dabei vorherrschend sind. Unbestritten scheint hingegen, dass Elicited Imitation sowohl das Kurz- als auch das Langzeitgedächtnis des Prüflings beansprucht: Die Testperson hört eine Äusserung, formt anhand ihres Langzeitvorwissens eine Darstellung und lagert sie in ihrem Kurzzeitgedächtnis ab. Anschliessend produziert sie einen Satz basierend auf der Darstellung, die sie kreiert hat (vgl. Graham et al. 2008, S. 1604). Aus diesem Grund wird nicht nur die mündliche Produktion, sondern auch das Verstehen in der Zweitsprache überprüft (vgl. Bowden 2016, S. 653).

Beim Elicited Imitation zählt das Prinzip: Je besser eine Fremdsprache beherrscht wird, desto besser kann sie auch imitiert werden. Elicited Imitation gilt daher als geeignete Testform, um allgemeines mündliches Sprachwissen zu messen (vgl. Bley-Vroman/Chaudron 1994, S. 247; Bowden 2016, S. 651–653). Da durch die Methode allgemeines mündliches Sprachwissen und demnach implizit grammatisches Wissen abgefragt wird, indem der Lernende Sätze möglichst genau wiedergeben soll, ist dieses Testinstrument auch bei Personen angebracht, die im Schreiben und Lesen weniger fortgeschritten sind oder noch (deutlich) Mühe damit haben (vgl. Tarone et al. 2009, S. 92; Wu/Ortega 2013, S. 684). Mit Hilfe der Elicited Imitation können weitere sprachliche Phänomene in der Fremdsprachenerwerbsforschung untersucht werden, wie beispielsweise der Wortschatz (vgl. Baten 2015, S. 172). Insgesamt lässt sich dieses Testinstrument sinnvoll für die Stichprobe dieser Studie anwenden, da Zweitschriftlernende in der Regel Schwierigkeiten mit dem deutschen bzw. dem lateinischen Schriftsystem haben.

Es gibt in der Zweitspracherwerbsforschung zahlreiche Studien, die Elicited Imitation für unterschiedliche Sprachen eingesetzt haben, so etwa für Englisch, Französisch oder Japanisch. Die Verwendung im Deutschen (vgl. Wu/Ortega 2013, S. 685) oder mit Zweitschriftlernenden und weniger schreibkundigen Personen ist hingegen bisher unerforscht.

Der Elicited Imitation von Ortega et al. (2002), der in diesem Kontext zunehmend an Bedeutung gewinnt, wurde ursprünglich für die Sprachen Englisch, Deutsch, Japanisch und Spanisch konzipiert. Das Ziel war, ein Set an Elicited Imitation in verschiedenen Sprachen zu entwickeln, die in einer sprachübergreifenden Studie zur Messung der Syntax benutzt werden sollten. Seit 2002 sind auf der Basis von Ortega et al. (2002) weitere Tests in anderen Sprachen erstellt worden, so z. B. in Mandarin (Wu/Ortega 2013) und in Französisch (Tracy-Ventura et al. 2014). Ortega et al. (2002) konnten der Elicited Imitation ein hohes Mass an Reliabilität, Konstruktvalidität sowie diagnostischer und prognostischer Validität nachweisen. Obwohl Ortega et al. im Jahr 2002 eine deutsche Elicited Imitation entwickelt haben (vgl. Ortega et al. 2002), konnte diese hier nicht verwendet werden, da die Sätze zum Teil ein höheres Deutschniveau voraussetzen, als es die Teilnehmenden der vorliegenden Studie aufwiesen.

Die häufigste Kritik an der Elicited Imitation besteht darin, dass nicht unterschieden werden kann, ob der Teilnehmende die Äusserung tatsächlich versteht oder lediglich deren Melodie imitiert, ohne dass der Sinn verstanden wird. Im ersten Fall kodiert die Person die Äusserung anhand ihrer grammatischen Vorkenntnisse. Im zweiten Fall wird lediglich der Rhythmus nachgeahmt, womit zwar die Wahrnehmungsmotorik überprüft werden kann, nicht aber die Fähigkeit, die Sprachstruktur zu kodieren (vgl. Vinther 2002, S. 55).

Aus diesem Grund wurde in dieser Studie nach jeder Äusserung eine 2.5 Sekunden lange Pause eingefügt. McDade et al. (1982) haben nachgewiesen, dass Partizipanten auch Äusserungen, die sie nicht verstehen, wiederholen können, wenn sie diese unmittelbar nachsprechen können. Erfolgt die Äusserung um 2 bis 3 Sekunden zeitlich versetzt, führt das dazu, dass sie nur korrekt wiederholt werden kann, wenn der Stimulus auch wirklich verstanden wurde. Die Pause ermöglicht daher, dass sowohl das Verständnis wie die Fähigkeit des Sprechens in der Zweitsprache erfasst werden kann (vgl. Bowden 2016, S. 653).

4.2.1.1 Testkonstrukt

Um die zuvor angehörte Äusserung zu wiederholen, muss die betreffende Sprachstruktur bereits als Interimssprache („Interlanguage“) beim Lernenden vorhanden sein (vgl. Erlam 2006, S. 467). Damit Lernende die Äusserung demzufolge wiederholen können, müssen sie diese auf

verschiedenen Ebenen verarbeiten und verstehen können. Die Prozesse, die während der Elicited Imitation im Detail ablaufen, können der unten stehenden Abbildung (s. Abb. 4) entnommen werden.

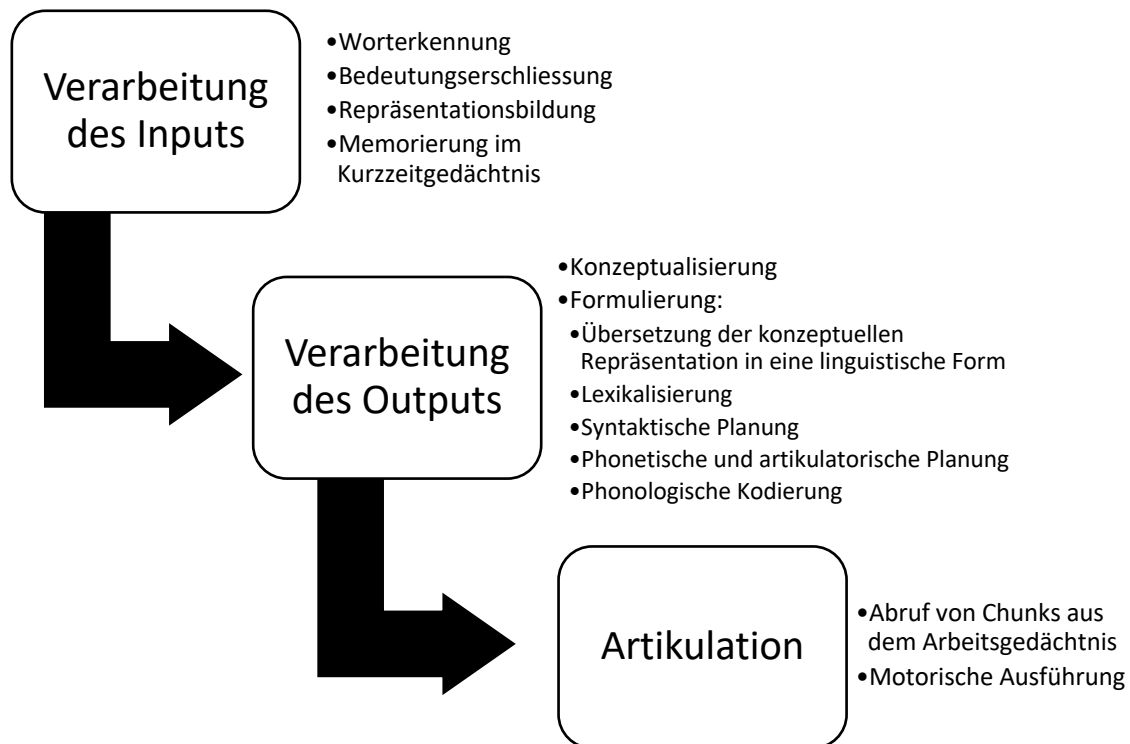


Abb. 4: Prozesse bei der Elicited Imitation (i. A. an Drackert 2015, S. 51)

Der Lerner muss gemäss Abb. 4 zunächst die Äusserung verarbeiten, verstehen und eine Darstellung formen, die im Kurzzeitgedächtnis gelagert wird. Im zweiten Schritt werden unterschiedliche Stufen durchlaufen: Es geht darum, dass der Lerner sein grammatisches Wissen auf das Verstandene anwendet. Für diesen Schritt werden, je nachdem wie fortgeschritten der Lernende ist, erhebliche Speichermengen des Arbeitsgedächtnisses benötigt. Mit zunehmender Sprachbeherrschung wird dieser Speicher immer weniger gefordert. Dementsprechend können fortgeschrittene Teilnehmende längere Äusserungen wiederholen als Lernende aus dem Anfängerbereich. In der letzten Stufe des obigen Modells produziert der Lerner die Äusserung (vgl. Drackert 2015, S. 51–52).

Damit nicht die Fähigkeit zum Wiederholen oder Imitieren der Äusserung getestet wird, wurde nach jeder gehörten Äusserung eine Pause von 2 Sekunden eingebaut und ein Signaltone von .05 Sekunden, nach dem das Nachsprechen erfolgen sollte. Mit der gehörten Äusserung, der zeitlichen Verzögerung von insgesamt 2.5 Sekunden und der Wiederholung der gehörten Äusserung durch den Prüfling erfasst der vorliegende Elicited Imitation die mündliche Produktion in der Zweitsprache Deutsch.

4.2.1.2 Aufgabenformat

Für die Erstellung der Stimuli wurde auf die häufigsten 2'000 Wörter der Häufigkeitsliste von Tschirner (2016) zurückgegriffen, die dem Wortschatzwissen auf dem Niveau A2 entsprechen. Da es sich um einen Zweitschriftlernerkurs handelt, bei dem das Niveau A1 bei Kursende erreicht sein sollte, wurde der Fokus auf die häufigsten 1'000 Wörter gelegt, die diesem Niveau entsprechen (vgl. Tschirner 2016, S. 3–5). Damit der Test für Lernende, die bereits einen Deutschkurs absolviert haben, anspruchsvoll bleibt, wurden neun von insgesamt 30 Items anhand der 2'000 häufigsten Wörter kreiert (A2-Niveau) – was insgesamt 30 % der Items entspricht. Dadurch konnte vermieden werden, dass Kursteilnehmer mit Vorkenntnissen die volle Punktzahl erreichen und eine Unterscheidung im oberen Quartil nicht mehr möglich ist. Die Wörter wurden mit Hilfe eines Zufallsgenerators ausgewählt und anschliessend wurden Sätze mit einer Silbenanzahl zwischen 4 und 16 kreiert. Obwohl sich in Studien zur Zweitspracherwerbsforschung eine Silbenlänge zwischen 6 und 19 Silben etabliert hat (vgl. Graham et al. 2010, S. 59), wurde in diesem Forschungsprojekt mit kürzeren Äusserungen gearbeitet. Ein Grossteil der bisherigen Forschung wurde mit bereits erfahreneren Lernenden durchgeführt, sodass die Silbenlänge hier an die teilweise geringeren Kenntnisse der Zielgruppe angepasst wurde. Unterschiedliche Stimuli-Längen wurden benutzt,

[...] because memory limitations are crucially involved, we expect accuracy when length is short. As length increases, accuracy will remain good until the limits of memory are approached. Then accuracy should fall quickly and remain. Around the limits of memory, there should be a narrow band of sensitivity, where accuracy might be affected by details of the syntactic structure (Bley-Vroman/Chaudron 1994, S. 252).

Laut Bley-Vroman und Chaudron ist somit anzunehmen, dass Anfänger kurze Äusserungen korrekt wiederholen können und ab einer bestimmten Silbenlänge durchgehend Mühe mit der Imitation haben. In den unten stehenden Tabellen (s. Tab. 9, Tab. 10 und Tab. 11) ist das Gesamtspektrum der Silben pro Messzeitpunkt tabellarisch dargestellt, aufgeschlüsselt nach Anzahl Items und Wörter der Häufigkeitsliste.

Häufigkeit der Wörter	Äusserungslänge in Silben												
	4	5	6	7	8	9	10	11	12	13	14	15	16
0–1000 (A1)	2	1	3	2	2	2	1	3	1	1	1	1	1
1000–2000 (A2)		1	1	2			1	1			1		2

Tab. 9: Äusserungslänge Elicited Imitation, Einstufungstest

Häufigkeit der Wörter	Äusserungslänge in Silben												
	4	5	6	7	8	9	10	11	12	13	14	15	16
0–1000 (A1)		2	4	2	3	3	1	3		2		1	
1000–2000 (A2)		2		1	2	1	1	1			1		

Tab. 10: Äusserungslänge Elicited Imitation, Follow-up 1

Häufigkeit der Wörter	Äusserungslänge in Silben												
	4	5	6	7	8	9	10	11	12	13	14	15	16
0–1000 (A1)		1	2	3	4	3	1	2			2	1	2
1000–2000 (A2)				1	2		1	1			3		1

Tab. 11: Äusserungslänge Elicited Imitation, Follow-up 2

Bei der Zusammenstellung wurde darauf geachtet, sowohl Aussage- als auch Frage- und Befehlsätze in den Äusserungen zu verwenden, wie dies auch in gängigen Deutschlehrwerken auf Niveau A1 der Fall ist. Die Verteilung der 30 Äusserungen im Einstufungstest stellt sich wie folgt dar: 50 % Aussagesätze, da diese den grössten Teil des Anfängerunterrichts ausmachen, je 20 % Frage- und Befehlssätze und 10 % Aussagesätze in der Vergangenheitsform. Perfekt und Präteritum werden üblicherweise erst zum Schluss des A1-Niveaus gelernt, deshalb wurden nur wenige entsprechende Äusserungen erstellt.

Die Sätze wurden alle durch die Forschende selbst und soweit möglich im Hinblick auf den Alltag der Teilnehmenden kreiert. Dafür wurden zuerst Wörter per Zufallsgenerator aus den Häufigkeitslisten von Tschirner (2016) herausgezogen. Anhand des ausgewählten Wortes wurde anschliessend ein Satz mit einer Silbenzahl zwischen 4 und 16 gebildet. Dabei wurde darauf geachtet, dass alle benutzten Wörter aus dem kreierten Satz zur Gruppe der 1'000 häufigsten Wörtern gehörten, um das Niveau A1 zu gewährleisten. Einzelne Sätze beziehen sich daher nicht unbedingt auf den Alltag der Teilnehmenden, beispielsweise *Die neue Küche ist sehr dunkel* (Follow-up 2, Äusserung 10). Die Sätze wurden anschliessend durch eine erfahrene Deutschkursleiterin überprüft, die die Möglichkeit hatte, Verbesserungen vorzuschlagen. Ihr wurden die Äusserungen nach einer ersten Überarbeitung nochmals vorgelegt. Der Prozess der Äusserungskreation wurde in Anlehnung an Graham et al. (2010, S. 62) durchgeführt und hielt sich an folgende Kriterien:

1. Sentences had to be comprehensible and plausible.
2. The criterion of sentence length in syllables had to be followed.
3. The criterion of selecting items of a given lexical frequency had to be followed. [...].

Im Anschluss daran wurden die Items erprobt (s. Abschn. 5.1), damit sie anschliessend jeweils durch eine weibliche und männliche Person im Tonstudio aufgenommen werden konnten, um eine standardisierte Durchführung zu gewährleisten.

Die Tonbandaufnahmen begannen stets mit der Instruktion, gefolgt von zwei Beispielen inkl. Wiedergabe eines bereits aufgenommenen Nachsprechens und weiteren drei Beispielen zum Wiederholen durch den Prüfling. Die Lernenden hörten bei der Testdurchführung alternierend eine Äusserung, die durch eine weibliche und eine, die durch eine männliche Stimme gesprochen wurde. Im Anschluss an jede Äusserung folgte eine Pause von 2 Sekunden und

ein .05 Sekunden dauernder Signalton, nach dem mit dem Nachsprechen begonnen werden konnte, d. h. zwischen den Äusserungen wurde für 2.5 Sekunden pausiert. Die Teilnehmenden hatten anschliessend 12 Sekunden Zeit, um die Äusserung möglichst genau zu wiederholen, wobei die Reaktionszeit an Graham (2010, S. 63) orientiert ist. Ein weiterer Signalton kündigte jeweils die nächste Äusserung an. Der gesamte Elicited Imitation wurde mit einem Aufnahmegerät aufgezeichnet und dauerte zu allen Messzeitpunkten knappe 13 Minuten inkl. Instruktion und fünf Beispielen. Die benutzten Äusserungen an den jeweiligen Messzeitpunkten sind aus den unten stehenden Tabellen (s. Tab. 12, Tab. 13 und Tab. 14) ersichtlich, wobei folgende neun Äusserungen aus dem Einstufungstest als Retest-Items für beide Follow-up-Messungen ausgewählt wurden (s. Abschn. 6.1.2.2):

- E8: Such das im Internet!
- E9: Muss er der Strasse folgen?
- E12: Ich habe dich getroffen.
- E14: Die Kinder helfen der Mutter.
- E15: Lernen Sie wenigstens fünf Wörter!
- E19: Meine Töchter sind unterschiedlich gross.
- E20: Haben Sie einen guten Eindruck von mir?
- E21: Beginnt deine neue Arbeit im August? (A2-Niveau)
- E26: Ich denke jeden Tag an unsere Vergangenheit. (A2-Niveau)

Das heisst, im Einstufungstest und den beiden Folgemessungen wurden je 21 Stimuli geprüft, die nur zu diesem Messzeitpunkt benutzt wurden, während neun weitere Stimuli, die Retest-Items, zu jedem der drei Testzeitpunkte verwendet wurden. Pro Messzeitpunkt wurden demnach 30 Stimuli getestet.

	Äusserung	Silbenzahl	Niveau
1.	Passiert das oft?	4	A1
2.	Der Zug hält hier.	4	A1
3.	Gib mir etwas Fisch !	5	A2
4.	Das kommt manchmal vor.	5	A1
5.	Gib mir eine Chance!	6	A1
6.	Sie sind enge Freunde.	6	A1
7.	Ich esse gerne scharf .	6	A2
8.	Such das im Internet!	6	A1
9.	Muss er der Strasse folgen?	7	A1
10.	Kauf kein teures Telefon !	7	A2
11.	Ich habe dich getroffen.	7	A1
12.	Mein Mann hat kochen gelernt.	7	A2
13.	Lies die Bücher heute Abend!	8	A1
14.	Die Kinder helfen der Mutter.	8	A1
15.	Lernen Sie wenigstens fünf Wörter!	9	A1
16.	Ich habe kein Wochenende frei.	9	A1
17.	Hat die Polizei Verständnis dafür?	10	A2
18.	Meine Töchter sind unterschiedlich gross.	10	A1
19.	Haben Sie einen guten Eindruck von mir?	11	A1
20.	Beginnt deine neue Arbeit im August ?	11	A2
21.	Ich verspreche, dass ich nach Hause komme.	11	A1
22.	Kannst du an die Universität gehen?	11	A1
23.	Meine Schwester macht Sport, weil sie keine Kraft hat.	12	A1
24.	Ich habe ein Mittel gegen die Krankheit gekauft.	13	A1
25.	Das Geschäft hat internationale Mitarbeiter.	14	A1
26.	Ich denke jeden Tag an unsere Vergangenheit .	14	A2
27.	Deine Schwester kann mir morgen auf die Frage antworten.	15	A1
28.	Bitte an die unten angegebene Adresse schicken.	16	A2
29.	Mein Bruder muss jeden Tag wegen einer Krankheit zum Doktor.	16	A1
30.	Jeden Donnerstag haben viele Geschäfte lange offen.	16	A2

Tab. 12: Äusserungen Elicited Imitation, Einstufungstest
Kursive Ziffer: Retest-Item; fettgedruckt: A2-Wort

	Äusserung	Silbenzahl	Niveau
1.	Ist die Schule nah?	5	A1
2.	Ist dein Büro gross?	5	A2
3.	Das Dorf liegt zentral.	5	A1
4.	Das Hotel ist schön.	5	A2
5.	Ich verstehe kein Deutsch.	6	A1
6.	Jeden Montag ist Markt.	6	A1
7.	Melden Sie sich täglich!	6	A1
8.	Findest du Sport auch ganz toll?	7	A1
9.	Ich benötige sofort Geld.	8	A1
10.	Meine Kinder haben oft Streit .	8	A2
11.	Rede mit deinem Kollegen!	8	A1
12.	Besteh die praktische Prüfung !	8	A2
13.	Ist das besonders interessant?	9	A1
14.	Er stellt das Gerät in die Küche.	9	A1
15.	Ich wohne in der Hauptstadt der Schweiz.	9	A2
16.	Ich habe hundert Euro gewechselt.	10	A1
17.	Er hat ein grosses Stück Fleisch gegessen.	10	A2
18.	Hänge die Bilder gerade an die Wand!	11	A1
19.	Sie möchte an der Universität studieren.	13	A1
20.	Alle Kinder wollen heute gemeinsam essen.	13	A1
21.	In der Nacht kann ich nur schlecht schlafen, weil es zu dunkel ist.	15	A1

Tab. 13: Äusserungen Elicited Imitation, Follow-up 1
fettgedruckt: A2-Wort

	Äusserung	Silbenzahl	Niveau
1.	Der Zeitpunkt ist schlecht.	5	A1
2.	Der Sommer beginnt spät.	6	A1
3.	Schreiben Sie beide Briefe!	7	A1
4.	Die Sonne scheint plötzlich stark.	7	A1
5.	Der Kollege überzeugt mich.	8	A1
6.	Zeigen Sie das rechte Auge!	8	A1
7.	Meine Frau hat ein Bett gekauft.	8	A1
8.	Ich habe perfekt geschlafen.	8	A2
9.	Mein Fortschritt im Schreiben ist gross.	8	A2
10.	Die neue Küche ist sehr dunkel.	9	A1
11.	Wo genau befindet sich das Dorf?	9	A1
12.	Finden die Schüler die Aufgabe leicht?	10	A1
13.	Was macht deine Schwester in der Freizeit ?	10	A2
14.	Hat seine Familie viele Tiere zu Hause?	14	A1
15.	Ich habe keine Zeit, denn ich habe einen Termin .	14	A2
16.	Unterstütze deine Frau sofort bei der Aufgabe!	14	A1
17.	Verabschiede dich heute Morgen von deiner Mutter!	14	A2
18.	Alle Informationen und Daten stehen im Brief.	15	A1
19.	Meine Familie interessiert sich für viele Dinge.	16	A1
20.	Ich sehe jeden Tag viele junge Menschen auf der Strasse.	16	A1
21.	Der Unterricht ist interessant und ich lerne immer viel.	16	A2

Tab. 14: Äusserungen Elicited Imitation, Follow-up 2
fettgedruckt: A2-Wort

4.2.1.3 Bewertungsverfahren

Alle Äusserungen der Teilnehmenden wurden mit Hilfe eines Audiogeräts aufgenommen und anschliessend durch die Forscherin transkribiert. Dadurch sollte die nachfolgende Bewertung durch externe Personen vereinfacht werden: So waren von diesen weniger als 100 Seiten Transkription zu beurteilen, während andernfalls fast 40 Stunden Tonbandaufnahmen anzuhören gewesen wären.

Das Bewertungsraster (s. Tab. 15) wurde von Ortega et al. (2002) übernommen: Ein Prüfling konnte maximal 4 Punkte pro Äusserung erhalten und somit total 120 Punkte im Elicited Imitation erzielen.

Punkte	Verständnis
4	Perfekte Wiedergabe.
3	Korrekte inhaltliche Wiedergabe mit (un-)grammatischen Änderungen.
2	Inhaltliche oder formbezogene Änderungen die den Inhalt verändern.
1	Die Hälfte (oder weniger) der Äusserung wird wiederholt.
0	Schweigen oder unverständliche Wiederholung.

Tab. 15: Bewertungsraster Elicited Imitation (vgl. Ortega et al. 2002, S. 2)

Die Bewertung soll anhand eines Beispiels erklärt werden. Die Lernenden hörten, wie weiter oben beschrieben, einen Stimulus, den sie so genau wie möglich wiederholen mussten, beispielsweise:

„Schreiben Sie beide Briefe!“ (Follow-up 2, Stimulus 4)

Die Wiederholungen dazu fielen, wie zu erwarten, sehr unterschiedlich aus. Die volle Punktzahl, d. h. vier Punkte, wurde bei komplett korrekter Wiedergabe des Stimulus vergeben.

Die Äusserung ‚Schreiben die beide Briefe‘ wurde wegen grammatikalischer Änderung mit drei Punkten bewertet. Wiederholungen, die sich inhaltlich vom Stimulus unterschieden, wie etwa ‚Schreiben sie beide‘, erhielten zwei Punkte. Einen Punkt gab es, wenn weniger als die Hälfte der gehörten Äusserung wiederholt wurde, beispielsweise bei ‚Schneiden beine Briefe‘. Unverständliche Äusserungen wurden mit null Punkten bewertet. Bei unterschiedlicher Beurteilung durch die externe Bewerterin und die Forscherin wurde eine dritte Person beigezogen und gemäss der Mehrheitsentscheidung bewertet.

4.2.2 Kunstwörter

Beim Test mit den sogenannten Kunstwörtern („nonword repetition“) werden Lernende aufgefordert, bedeutungsfreie Wörter, die Ähnlichkeiten mit ‚richtigen‘ Wörtern und zudem eine bestimmte Anzahl an Silben aufweisen, nachzusprechen (vgl. Hasselhorn et al. 2000, S. 119). Dabei variieren die Kunstwörter in der Länge zwischen zwei und sechs Silben (vgl. Hasselhorn/Körner 1997, S. 213). Forscher betrachten das Nachsprechen von Kunstwörtern als Indikator für die Funktionstüchtigkeit des (phonologischen) Arbeitsgedächtnisses (vgl. Gathercole et al. 1994, S. 104–105; Hasselhorn et al. 2000, S. 213–214). Die Arbeitsgedächtniskapazität gilt als «im Vergleich zum Intelligenzquotient besseren [sic] Prädiktor schulischer Leistungen» (Seitz-Stein et al. 2012, S. 6).

Nach dem ursprünglichen Modell des Arbeitsgedächtnisses von Baddeley und Hitch (1974) ist dieses in drei Teile gegliedert:

die phonologische Schleife, zum Umgang mit verbaler Information, den visuell-räumlichen Skizzenblock, der die Manipulation von bildhaften und räumlichen Informationen unterstützt, und eine diesen beiden Hilfssystemen übergeordnete zentrale Exekutive, die als Steuer- und Kontrollinstanz des gesamten Systems konzipiert [...] ist (Baddeley/Hitch 1974, S. 3).

Die phonologische Schleife, die sich in einen phonologischen Kurzzeitspeicher und ein Rehearsal-System (Wiederholungs- bzw. Auffrischungssystem) differenzieren lässt, ist laut Baddeley und Hitch ein Mass für die Effizienz und Genauigkeit des phonologischen Kurzzeitspeichers. In diesen müssen die Kunstwörter beim Hören gespeichert werden, weil dazu keine Repräsentationen im mentalen Lexikon existieren. Da der phonologische Kurzzeitspeicher jedoch lediglich während eines Zeitfensters von zwei Sekunden passiv speichern kann, muss das Rehearsal-System einspringen, indem der Stimulus mittels phonologischer Rehearsal-Prozesse aktiv aufrechterhalten wird (vgl. Mayer 2016, S. 72).

Diese phonologische Schleife ist bis dato das am besten untersuchte Subsystem des Modells von Baddeley und Hitch und damit empirisch gut belegt (vgl. Seitz-Stein et al. 2012, S. 3–4). Im deutschsprachigen Raum zeigen zahlreiche Forschungsergebnisse, dass Kinder mit Lese-

Rechtschreibschwierigkeiten in Arbeitsgedächtnistests unterdurchschnittlich abschneiden und dass das phonologische Arbeitsgedächtnis einen signifikanten Prädiktor für spätere Leseleistungen darstellt (vgl. Mayer 2016, S. 74).

Bei Erwachsenen ist das Nachsprechen von Kunstwörtern wesentlich mit der Fähigkeit verbunden, neue phonologische Formen zu lernen, die keine Ähnlichkeit zu vertrauten Wörtern haben. Ältere Erwachsene scheinen damit mehr Mühe zu haben als jüngere Menschen (vgl. Gupta 2003; Gathercole 2006, S. 516). Castro-Caldas et al. (1998) konnten in ihrer Studie mit primären Analphabeten und schriffterfahrenen Personen zeigen, dass erstere beim Nachsprechen von Kunstwörtern schlechter abschneiden als schriffterfahrene Personen. So versuchten primäre Analphabeten, die Kunstwörter in Realwörter umzuwandeln. Laut Castro-Caldas et al. deutet dies darauf hin, dass die phonologische Sprachverarbeitung bei primären Analphabeten eine andere Kapazität aufweist als bei schriffterfahrenen Personen (vgl. ebd., S. 1060). Diese Überlegungen könnten eine Erklärung für ein schlechteres Abschneiden im Elicited Imitation sein. Das heisst, unterdurchschnittliche Ergebnisse im Elicited Imitation könnten auf ein ungenügend ausgebildetes Arbeitsgedächtnis zurückgeführt werden.

Das Arbeitsgedächtnis wird indes nicht nur für das Lesen und Schreiben benötigt, sondern spielt in vielen Bereichen des Alltags eine Rolle. So wurde die Bedeutsamkeit der phonologischen Schleife für den Wortschatzerwerb, das Sprachverständnis, das Lesenlernen, das Textverstehen und die Fähigkeit zur Durchführung von Rechenoperationen wie Addition und Multiplikation bestätigt (vgl. z. B. Hasselhorn et al. 2000, S. 120; Seitz-Stein et al. 2012, S. 4; Ulrich 2016, S. 23). Die Erfassung des Arbeitsgedächtnisses ist daher nicht nur eine geeignete Möglichkeit, um sprachliche Kompetenzen vorherzusagen, sondern spiegelt auch generell das Niveau der Grundbildung der Geprüften. Entsprechend hat der Arbeitsspeicher in den vergangenen Jahren zunehmende Beachtung auch in der Zweitspracherwerbsforschung erhalten. Aus diesen Studien ist ersichtlich, dass Lernende mit einem grösseren Arbeitsspeicher die Zweitsprache besser lernen und anwenden können (vgl. z. B. Wagner 2014, S. 51).

Ein Kritikpunkt beim Nachsprechen von Kunstwörtern und so auch beim Mottier-Test (s. Abschn. 4.2.2) bezieht sich auf die plosiven Laute, z. B. [k], sowie auf die schnellen, wechselhaften Artikulationsbewegungen, z. B. [ka pe to]. Diese können bei Personen mit eingeschränkten artikulationsmotorischen Fähigkeiten erhöhte Anforderungen an die

phonetischen und phonologischen Kompetenzen stellen (vgl. Ulrich 2016, S. 23), weshalb die Kapazität ihres Arbeitsgedächtnis unterschätzt werden kann. Diese zusätzliche Schwierigkeit wurde in Bezug auf den Mottier-Test jedoch noch nicht empirisch untersucht.

4.2.2.1 Testkonstrukt

Der Mottier-Test wurde von der Zürcher Logopädin Grete Mottier 1951 als Zusatz zum Zürcher Lesetest entwickelt und wird in der Sprachdiagnostik seither vielfach eingesetzt. Der Mottier-Test wird üblicherweise bei Kindern und Jugendlichen im Alter zwischen 4 und 17 Jahren verwendet, um Lese- und Rechtschreibstörungen zu erkennen (vgl. Wild/Fleck 2013, S. 153). Beim Mottier-Test werden die einzelnen Silben im Sekundentakt vorgesprochen. Der phonologische Kurzzeitspeicher reicht aber für das Nachsprechen der Kunstwörter nicht aus, da er, wie bereits dargestellt, lediglich ein Zeitfenster von 2 Sekunden umfasst und die Kunstwörter länger sind. Um die Kunstwörter nachsprechen zu können muss somit nebst dem phonologischen Kurzzeitspeicher auf das Rehearsal-System zurückgegriffen werden. Dank dieses Rückgriffs kann mit dem Mottier-Test die Gesamtkapazität der phonologischen Schleife erfasst werden, welche eine Komponente des Arbeitsgedächtnisses ist (vgl. Mayer 2016, S. 70–73). Bis anhin muss das Arbeitsgedächtnis bei Zweitschriftlernenden als unerforschter Bereich gelten.

Der Mottier-Test soll in der vorliegenden Forschung einerseits Unterschiede in der Kapazität des Arbeitsgedächtnisses der Teilnehmenden aufdecken. Andererseits werden die Kunstwörter mit den anderen Tests auf Korrelationen überprüft. So könnte sich ein schlechteres Abschneiden bei der Elicited Imitation, im Hörverstehen und/oder im Diktat auf das Arbeitsgedächtnis zurückführen lassen, da in allen diesen Tests das Arbeitsgedächtnis eine erhebliche Rolle spielt (s. Abschn. 7.3.3).

4.2.2.2 Aufgabenformat

Der Mottier-Test besteht aus insgesamt 30 zwei- bis sechssilbigen Kunstwörtern, wobei die Silben allesamt sinnleere Silbensequenzen aus Konsonant-Vokal-Verbindungen sind (s. Tab. 16). Diese Silbenfolge gilt es, in korrekter Reihenfolge unmittelbar nach dem Stimulus zu reproduzieren. Dabei ist anzumerken, dass sowohl plosive Laute als auch schnelle, wechselhafte Artikulationsbewegungen im Test vorkommen.

Mottier hat empfohlen, dass die Testleitenden die Wörter im Sekundentakt mit verdecktem Mundbild und gleichmässiger Betonung vorsprechen (vgl. Ulrich 2016, S. 22–23). Spätere Forschungen haben jedoch gezeigt, dass das Vorsprechen mit verdeckten oder sichtbaren

Mundbewegungen nicht zu signifikanten Unterschieden führt (vgl. Petermann/Daseking 2018, S. 70–71). Um eine standardisierte Wiedergabe der Kunstwörter zu ermöglichen, wurde auf eine dritte Variante zurückgegriffen: die Wiedergabe über ein Tonband. Hierzu wurden die Kunstwörter des Mottier-Tests vorgängig durch eine weibliche Sprecherin in einem Tonstudio aufgenommen. Dies geschah in Anlehnung an den Originaltest (vgl. Gamper et al. 2012, S. 4–5) im Sekundentakt und mit gleichbleibender Betonung. Die aufgenommenen Kunstwörter wurden anschliessend mit Hilfe einer entsprechenden Software („Audacity“) zu einer Datei geschnitten. Die Prüflinge hörten bei der Testdurchführung die Kunstwörter vom Band, wobei zuvor ein Beispiel inkl. Nachsprechen der Sprecherin und zwei weitere Beispiele zum Nachsprechen durch den Prüfling selbst eingespielt wurden. Anschliessend wurde mit dem eigentlichen Test begonnen, bei dem die Audiodatei ohne Unterbrechung abgespielt wurde: Die Teilnehmenden hörten das Kunstwort gefolgt von einem kurzen Signalton und hatten anschliessend 12 Sekunden Zeit, um das Kunstwort nachzusprechen. Danach ertönte ein etwas längerer Signalton, der das nächste Kunstwort ankündigte. Wurde eine ganze Spalte, bestehend aus sechs Kunstwörtern, falsch nachgesprochen, wurde der Test, wie im Originaltest vorgesehen, beendet. Die Durchführung des kompletten Tests dauerte circa 9 Minuten und wurde zu jedem Messzeitpunkt wiederholt. Der Mottier-Test wurde direkt nach dem Elicited Imitation durchgeführt und die Antworten der Teilnehmenden wurden mit einem Aufnahmegerät aufgezeichnet.

zweisilbig	dreisilbig	viersilbig	fünfsilbig	sechssilbig
re la	ka pe to	pi ka tu ra	ka to pi na fe	pe ka to ri se ma
no ma	gi bo da	ga bo di la	ge bi da fi no	da go bi la se ta
go du	lo re ma	mo na lu ra	ro na me li ta	le ra mi no fe ko
me ra	to ki pa	to pa ki mu	ta pi ku sa we	ka po ti la fe sa
lu ri	du ga be	de ba gu si	de go be sa ro	bi ga do na fe ra
li mo	no ma ri	re lo ma no	mu ra le no ka	no ma li ra ko sa

Tab. 16: Kunstwörter Mottier-Test (Gamper et al. 2012, S. 5)

4.2.2.3 Bewertungsverfahren

Die Erstbewertung des Mottier-Tests wurde während der Testdurchführung durch den jeweiligen Testleitenden mit Hilfe des Bewertungsbogens vorgenommen (s. Tab. 17). In einem zweiten Schritt wurden die ausgefüllten Bewertungsraster anhand der aufgenommenen Audiodateien durch die Forscherin überprüft und gegebenenfalls angepasst. Diese Überprüfung diente dazu, die durch verschiedene Personen ausgefüllten Prüfraster zu vereinheitlichen. So zeigte sich besonders bei der Beurteilung der Plosiv-Laut eine deutliche Diskrepanz zum Ergebnis der Erstprüfenden: Das dritte Kunstwort [go du] wurde hier z. B. mehrfach als korrekt angegeben, obwohl viele der Prüflinge [go tu] wiederholten.

Wie aus Tab. 17 zu entnehmen ist, konnten die Testleiter die Bewertung ‚richtig‘ ankreuzen, wenn das Kunstwort korrekt nachgesprochen worden war. In jedem Fall schrieb der Testleiter das gehörte Wort zusätzlich auf die vorgegebene Linie. Dieses Vorgehen ist nach der Pilotierung übernommen worden, bei der nur die falsch nachgesprochenen Wörter aufgeschrieben wurden. Da dieses Vorgehen die Teilnehmenden irritiert hat, wurde in der Hauptuntersuchung das Wort sowohl bei falscher als auch bei richtiger Wiederholung aufgeschrieben (s. Abschn. 5.1.1).

Wurde eine ganze Spalte à 6 Kunstwörter nicht korrekt wiederholt, wurde der Test beendet (vgl. Gamper et al. 2012, S. 4–5). Pro richtig gesprochenes Kunstwort wurde jeweils 1 Punkt vergeben, was ein Total von 30 Punkten ergibt.

zweisilbig		dreisilbig		viersilbig		fünfsilbig		sechssilbig	
re la		ka pe to		pi ka tu ra		ka to pi na fe		pe ka to ri se ma	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____
no ma		gi bo da		ga bo di la		ge bi da fi no		da go bi la se ta	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____
go du		lo re ma		mo na lu ra		ro na me li ta		le ra mi no fe ko	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____
me ra		to ki pa		to pa ki mu		ta pi ku sa we		ka po ti la fe sa	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____
lu ri		du ga be		de ba gu si		de go be sa ro		bi ga do na fe ra	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____
li mo		no ma ri		re lo ma no		mu ra le no ka		no ma li ra ko sa	
richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____	richtig <input type="checkbox"/>	falsch _____

Tab. 17: Bewertungsbogen Kunstwörter

4.2.3 Oral Reading Fluency

Oral Reading Fluency (ORF) ist die Bezeichnung für eine komplexe Leistung, die verschiedene Aspekte des Lesens umfasst, sowie ein Testinstrument, um die entsprechenden Prozesse zu messen (vgl. Fuchs et al. 2001, S. 239–240; Pretorius/Spaull 2016, S. 1452). Fuchs et al. (2001, S. 239–240) haben die Vielfältigkeit in ihrem Statement wie folgt zusammengefasst:

[...] oral reading fluency represents a complicated, multifaceted performance that entails, for example, a reader's perceptual skill at automatically translating letters into coherent sound representations, unitizing those sound components into recognizable wholes and automatically accessing lexical representations, processing meaningful connections within and between sentences, relating text meaning to prior information, and making inferences to supply missing information. That is, as an individual translates text into spoken language, he or she quickly coordinates these skills in an obligatory and seemingly effortless manner, and because oral reading fluency reflects this complex orchestration, it can be used in an elegant and reliable way to characterize reading expertise.

In der gängigen Literatur werden hauptsächlich drei Komponenten beschrieben, die anhand der ORF gemessen werden können: die Wortlesegenauigkeit, die Leseflüssigkeit und die Prosodie (vgl. Pretorius/Spaull 2016, S. 1454; Ciuffo et al. 2017, S. 1668). Fuchs et al. betrachten das Testinstrument der ORF als elegante und zuverlässige Möglichkeit, um Lesekompetenzen zu messen (vgl. Fuchs et al. 2001, S. 240). Das Vorgehen zeigt die Fähigkeit eines Lesers zum schnellen Koordinieren von verschiedenen Lesefähigkeiten (vgl. Mellard et al. 2012, S. 1346). Zudem ist ORF ein zuverlässiger Indikator für das Textverstehen, was in verschiedenen Studien ³¹ mit sogenannten ‚High-Stake‘-Tests und verschiedenen Testformaten (multiple choice, offene Fragen, ‚story recall‘ Protokollen etc.) belegt werden konnte (vgl. Pretorius/Spaull 2016, S. 1454).

Die ORF hat in den letzten Jahren in der Forschungslandschaft vermehrt Beachtung gefunden, um die allgemeine Lesefertigkeit hauptsächlich von Primarschülern und Jugendlichen in der L1 und in geringerem Masse auch in der L2 zu untersuchen (vgl. Petscher/Kim 2011, S. 108). So konnte in älteren Studien (s. z. B. Fuchs et al. 2001, S. 243–248) eine deutliche Korrelation zwischen mündlicher Leseflüssigkeit und Textverstehen bei Grundschulkindern und Jugendlichen gezeigt werden. In der Leseerwerbsforschung bei Primarschulkindern wird die ORF für die Lesediagnostik oder für das Messen der allgemeinen Lesefortschritte benutzt (s. z. B. Mellard et al. 2011; Mellard et al. 2012; Gagliano et al. 2015). In Alphabetisierungskursen mit Erwachsenen wird ORF als sogenanntes Lautleseprotokoll verwendet, das entweder als förderdiagnostisches Instrument oder wie bei Kindern zur Fortschrittsmessung dient (vgl. Scherf/Rieckmann 2016).

Mit dem Verfahren, nach dem die korrekt gelesenen Wörter pro Minute gezählt werden, auch curriculum-based measurement (CBM) genannt, können die Leistungen der einzelnen Teilnehmenden verglichen sowie Unterschiede in der Entwicklung der Lesefähigkeit individuell untersucht werden. CBM kann folglich auch als eine Art Förderinstrument dienen, da Lehrpersonen beispielsweise mit Lautleseprotokollen Fehleranalysen erstellen können. Aus diesem Grund wird die Methode des CBM auch direkt durch Lehrpersonen angewendet und ausgewertet (vgl. Fuchs et al. 2001, S. 240, 251–252), damit frühzeitig festgestellt werden kann, welche Lernenden besonderen Förderbedarf aufweisen (vgl. Pretorius/Spaull 2016, S. 1455).

³¹ Für einen Überblick über diese Forschungen siehe Fuchs et al. (2001, S. 243–248).

Da noch wenige Forschungen zu ORF vorliegen, existieren bisher keine Schwellenwerte für die Leseflüssigkeit im Deutschen, weder für Kinder noch für Erwachsene. Aufgrund fehlender empirischer Untersuchungen lässt sich zudem nicht beantworten, inwiefern sich die Textsorte auf die Korrelation von ORF und Textverstehen auswirkt (vgl. Fuchs et al. 2001, S. 252). Da in der vorliegenden Studie die ORF nicht als förderdiagnostischer Test betrachtet wird und keine Massnahmen aufgrund des Testergebnisses ergriffen werden sollen, können diese Kritikpunkte für diese Studie vernachlässigt werden. Bassetti et al. (2012, S. 6) halten ORF für eine nicht geeignete Methode, um Lesen in der L2 zu untersuchen, und nehmen dabei Bezug auf die Befunde bei Lems (2003). Lems (2003) konnte in ihrer Studie mit Zweitschriftlernenden ($N = 232$) jedoch eine hochsignifikante leichte bis moderate positive Korrelation zwischen ORF und Textverstehen zeigen (vgl. ebd., S. 116–125). Mellard et al. (2011) konnten in ihrer Studie mit Erwachsenen gar einen deutlichen Zusammenhang zwischen Leseflüssigkeit und Textverstehen nachweisen.

4.2.3.1 Testkonstrukt

Obwohl ORF mehrere Komponenten umschliesst, wird in dieser Studie die Leseflüssigkeit untersucht und nicht, wie in anderen Studien, die Wortlesegenauigkeit oder die Prosodie. Zu letzteren beiden gehören Komponenten wie: Lesegenauigkeit, Automatisierungsgrad des Dekodierens auf Wortebene, Lesegeschwindigkeit und prosodische Segmentierungsfähigkeit auf Satzebene (vgl. Fuchs et al. 2001, S. 247).

Die ORF vermag individuelle Unterschiede «in a number of reading subcomponents at lower and higher levels of processing» zu erfassen (ebd.). Die einzelnen Dimensionen sind als eng zusammenhängend zu betrachten (vgl. Rosebrock et al. 2017, S. 15–16), sodass die Leseflüssigkeit (‘reading fluency’) als «Fähigkeit zur genauen, automatisierten, schnellen und sinnkonstituierenden leisen und lauten Lektüre» (ebd., S. 15) als autonomer Bestandteil der Lesekompetenz verstanden werden kann. In anderen Worten ausgedrückt geht es darum, einen Text mühelos und routiniert zu lesen (vgl. ebd.). Damit gilt diese Komponente als Vorläuferfähigkeit des Textverstehens, denn laut den Autoren setzt erst ein gewisser Grad der Leseflüssigkeit auf der Wort- und Satzebene die kognitiven Ressourcen frei, die für höhere Verstehensprozesse beim Lesen benötigt werden. Entsprechend ist es nicht verwunderlich, dass ORF ein deutlicher Indikator für das Textverständnis bei Lernenden jeden Niveaus ist, besonders jedoch bei schwachen Lesern (s. z. B. Sturm 2011, S. 15; Petscher/Kim 2011, S. 109–110).

Die vorliegende ORF stellt keinen förderdiagnostischen Test im Sinne eines Lautleseprotokolls dar, sondern eine Erhebung der Leseflüssigkeit, gemessen anhand der Anzahl korrekt gelesener Wörter pro Minute (WpM). In dieser Studie werden die Ergebnisse der ORF, ähnlich wie bei Lems (2003), auf die Korrelation mit dem Textverstehen von telc geprüft.

4.2.3.2 Aufgabenformat

Bei der ORF ist es von zentraler Bedeutung, dass der ausgewählte Lesetext dem Niveau der Lernenden entspricht (vgl. ebd., S. 157). In dieser Forschungsarbeit bestanden in der Stichprobe unterschiedliche Vorkenntnisse, von absoluten Anfängern bis hin zu fortgeschrittenen Lernern im A-Niveaubereich. Um allen Teilnehmenden gerecht zu werden, wurde ein Lesetext, mit freundlicher Genehmigung des Verlags, aus dem Niveau A1 gewählt. Es wurde bewusst auf Lesetexte aus gängigen Lehrwerken verzichtet, damit ausgeschlossen werden konnte, dass Lernende, die bereits einen Deutschkurs besucht haben, diese kannten. Der Text *Ralf* aus dem *Lesebuch von A bis Z – Leseheft für den Alphabetisierungskurs Lesestufe 1* (Fellmer/Feldmeier 2012, S. 50–62) entspricht diesen Kriterien und weist eine ausreichende Textlänge auf. Der Text wurde in drei gleich grosse Teile aufgeteilt, für den Einsatz zu jedem Messzeitpunkt (s. Tab. 18, Tab. 19 und Tab. 20). Einige Sätze wurden ausgelassen, sodass der Lesbarkeitsindex (vgl. Lenhard/Lenhard 2017) eine leichte Progression aufwies. Der in der Einstufung verwendete Text weist einen Lesbarkeitsindex von 18.0 bei 173 Wörtern auf, der Text bei der ersten Follow-up-Messung einen Index von 18.2 bei 161 Wörtern und der Text bei der zweiten Follow-up-Messung einen Wert von 18.3 bei 168 Wörtern. Somit waren alle Lesbarkeitsindexe unterhalb von 20 angesiedelt, was sehr leichten Texten entspricht (vgl. Nickel 2011, S. 31). Da verschiedene Studien zeigen konnten, dass eine Minute eine hinreichende Zeitdauer darstellt, um die Leseflüssigkeit zu messen (s. z. B. Fuchs et al. 2001; Lems 2003, S. 95), wurde dies auch für diese Studie übernommen.

Im Anschluss an die Kunstwörter wurden die Teilnehmenden instruiert, den Text eine Minute lang so genau wie möglich laut zu lesen (vgl. Ciuffo et al. 2017, S. 1672). Die erste, bei allen Messungen gleichlautende Zeile zählte noch nicht zum eigentlichen Test, sondern sollte die Lernenden für das Lesen gewissermassen ‚aufwärmen‘ (vgl. Lems 2003, S. 42). Die Minute wurde ab dem Wort *Ich [lese gern]* gestoppt, wobei die ersten vier Sätze zu jedem Messzeitpunkt die gleichen waren. Während der Durchführung wurde die Produktion der Teilnehmenden akustisch aufgenommen, sodass die Bewertung im Anschluss durch die Forscherin erfolgen konnte.

Hallo! Mein Name ist Ralf. Ich bin Deutscher. Ich wohne in Frankfurt. Ich spiele gern Fussball.
Ich lese gern. Ich arbeite bei der Post. Die Arbeit macht mir Spass. Mein Leben macht mir heute viel Spass. Ich bin glücklich. Früher war ich nicht glücklich. Warum? Hier ist meine Geschichte. Ich bin sechs Jahre alt. Ich komme in die Grundschule. In meiner Klasse sind viele Kinder. Die Kinder lernen gut. Ich kann nicht so gut lernen. Ich kann nicht gut lesen und ich kann nicht gut schreiben. Die Lehrerin kann mir nicht immer helfen. Die Lehrerin muss auch den anderen Kindern helfen. Meine Eltern wollen mir helfen. Sie schaffen es auch nicht. Ich muss ein Schuljahr wiederholen. Ich komme in die Hauptschule. Die anderen Kinder wissen schon sehr viel. Ich weiss nicht so viel. Ich muss vorlesen. Ich kann nicht vorlesen. Es geht nicht. Was soll ich machen? Ich strenge mich noch mehr an. Ich will wie die anderen Kinder sein. Ich will auch gute Noten haben. Aber es geht nicht. Ich bin traurig. Ich habe Angst. Manchmal gehe ich nicht zur Schule. Ich muss noch mal ein Jahr wiederholen.

Tab. 18: ORF-Text, Einstufung (Fellmer/Feldmeier 2012, S. 50–62)

Hallo! Mein Name ist Ralf. Ich bin Deutscher. Ich wohne in Frankfurt. Ich spiele gern Fussball.
Ich lese gern. Ich arbeite bei der Post. Die Arbeit macht mir Spass. Ich bin 17. Ich muss einen Beruf lernen. Mein Vater kennt einen Mann mit einer Autowerkstatt. Ich darf dort eine Ausbildung machen. Ich habe Angst. Muss ich bei der Arbeit schreiben? Ich kann das nicht. Mein Chef und meine Kollegen wissen das nicht. Mein Kollege sagt: «Ralf, schreib mal.» Ich antworte: «Ich kann jetzt nicht. Meine Hände sind dreckig.» Oder: «Schreib du. Ich muss auf die Toilette.» Die Arbeit ist anstrengend. Lügen ist auch anstrengend. Ich bin so müde. Überall muss man lesen oder schreiben. Ich kann nicht lesen und schreiben, aber das darf kein Mensch wissen. Ich möchte das nicht. Ich schäme mich! Ich habe Angst. Ich verstecke mich. Ich mache viele Sachen nicht: Ich fahre nicht weg. Ich gehe in keinen Verein. Ich gehe nicht zum Arzt. Ich gehe nicht gern mit anderen Menschen essen. Ich lese keine SMS. Ich bin alleine. Ich muss etwas ändern.

Tab. 19: ORF-Text, Follow-up 1 (Fellmer/Feldmeier 2012, S. 50–62)

Hallo! Mein Name ist Ralf. Ich bin Deutscher. Ich wohne in Frankfurt. Ich spiele gern Fussball.
Ich lese gern. Ich arbeite bei der Post. Die Arbeit macht mir Spass. Ich habe von einer Schule für Erwachsene gehört. Ich rufe dort an. Die Frau von der Schule ist sehr nett. Sie sagt: «Kommen Sie nächsten Montag um 16 Uhr.» Ich bin sehr aufgeregt. Ich stehe vor der Schule. Die Kursleiterin und fünf Personen sind im Raum. Sie wollen alle lesen und schreiben lernen. Ich spreche mit den anderen Männern und Frauen. Alle hatten Angst und haben gedacht: Wieder Schule. Wieder lernen. Können wir das schaffen? Es ist schwer. Wir machen viele Übungen. Die Kursleiterin macht uns Mut. Sie hilft und hat Zeit für alle. Heute kann ich lesen und schreiben. Ich habe keine Angst mehr und ich verstecke mich nicht mehr. Ich mache jetzt viele Sachen: Ich fahre weg. Ich gehe in einen Verein. Ich gehe zum Arzt, wenn ich muss. Ich habe einen Computer. Ich lese Bücher. Und ich lese SMS. Ich helfe anderen Menschen. Ich bin glücklich. Alle können lesen und schreiben lernen!

Tab. 20: ORF-Text, Follow-up 2 (Fellmer/Feldmeier 2012, S. 50–62)

4.2.3.3 Bewertungsverfahren

Im Gegensatz zum Elicited Imitation wurden beim ORF nicht Transkriptionen bewertet, sondern es wurden die Audiodateien, insgesamt etwa über drei Stunden Material, durch die Forscherin direkt angehört und beurteilt. Die Bewertung erfolgte in Anlehnung an Pretorius/Spaull (2016, S. 973) und Gagliano et al. (2015, S. 1452) und wurde durch die Forscherin und einer Zweitbewerterin vorgenommen. Dazu wurden die falsch gelesenen Wörter von der Gesamtzahl gelesener Wörter subtrahiert, was ein Total korrekt gelesener Wörter pro Minute ergab und die Leseflüssigkeit der Prüflinge in der Zweitsprache Deutsch widerspiegelte. Als falsch gelesen gelten Wörter, wenn ein oder mehrere Buchstaben ausgelassen, hinzugefügt oder ausgetauscht wurden oder die Wiedergabe unverständlich war. Gängige Fehler waren beispielsweise ‚lese‘ statt ‚lesen‘, ‚gerne‘ statt ‚gern‘ oder ‚der‘ statt ‚die‘.

4.2.4 Hörverstehen

Obwohl Hörverstehen in der ersprachlichen Alltagskommunikation den ersten Platz unter den Fertigkeiten einnimmt – ungefähr 45 bis 50 % der kommunikativen Situationen bestehen aus Hören (vgl. Brinitzer et al. 2013, S. 24; Wagner 2014, S. 47) – wird es im Vergleich zu den anderen Fertigkeiten bei der Schulung sowie der Überprüfung in der Zweitsprache vernachlässigt (vgl. Grotjahn 2001, S. 3). Auch sind viele Lehrpersonen und Lernende der Ansicht, diese Fertigkeit würde sich von selbst weiterentwickeln (vgl. Flowerdew/Miller 2014, S. 225). Ein möglicher Grund für die geringe Beachtung des Hörens durch die Forschung könnte sein, dass Hörverstehen ein äusserst komplexer Vorgang ist (vgl. Wagner 2014, S. 47), bei dem «weder der Prozess des Verstehens noch das erzielte Verständnis unmittelbar beobachtbar ist» (Grotjahn 2001, S. 5). Es kann weder auf Studien zum Hörverstehen mit Zweitschriftlernenden noch auf solche mit Analphabeten zurückgegriffen werden.

Dem Lese- und Hörverstehen werden oft einander ähnliche kognitive Prozesse zugeschrieben, weil sie beide mit einem Input beginnen und in den Aufbau mentaler Konzeptualisierungen von Botschaften münden (vgl. Green 2014, S. 104). Im Gegensatz zum Leseverstehen ist Hörverstehen jedoch ein ‚Echtzeit-Prozess‘ (vgl. z. B. Porsch et al. 2010, S. 152), da Hören flüchtig ist und die Geschwindigkeit der Informationsübermittlung durch den Hörenden nicht beeinflusst werden kann (vgl. z. B. Nold/Henning 2007, S. 178; Porsch et al. 2010, S. 152). Hörverstehen umfasst sowohl Bottom-up-Prozesse, bei denen einzelne Informationen aus dem Text vom Hörer aufgenommen werden, als auch Top-down-Prozesse, durch die das Sach- und das Sprachwissen des Hörers an den Text herangetragen werden (vgl. Buck 2001, S. 40–42). Dadurch entsteht ein «mentales Abbild der gehörten Informationen» (Solmecke 2010, S. 970).

Texte für das Hörverständnis können leicht zu einer Überbelastung des Arbeitsgedächtnisses führen, da der Rezipient weder die Wiedergabegeschwindigkeit beeinflussen noch das Gehörte nach Belieben wiederholen lassen kann. Die übermittelten Informationen müssen deshalb alle im Arbeitsgedächtnis zwischengespeichert werden, was dadurch erschwert wird, dass der auditive Kanal eine geringere Verarbeitungs- und Speicherkapazität als beispielsweise der visuelle Kanal aufweist. Zudem handelt es sich bei einem Hörtext um kontinuierlich übermittelte akustische Signale, bei denen die Wortgrenzen nur teilweise markiert sind. Daher muss der Rezipient den kontinuierlichen Lautstrom zumindest teilweise in Einzelwörter segmentieren. Gerade bei fremdsprachlichen Texten kann dies eine grosse

Schwierigkeit darstellen, da hierzu ein gewisser Wortschatz nötig ist (vgl. Grotjahn 2001, S. 5–6).

Besonders Zweitsprachlernende haben Mühe im Hörverstehen, da das Wissen in der Zweitsprache in vielen Bereichen in der Regel unvollkommen ist (vgl. Nold/Henning 2007, S. 178–179): «When second language listeners are listening, there will often be gaps in their understanding. [...] Of course, gaps occur in first-language listening too, but the gaps in second-language listening usually have a far more significant impact on comprehension» (Buck 2001, S. 50).

4.2.4.1 Testkonstrukt

Hörverstehen in der Zweitsprache kann u. a. aus psycholinguistischem, kognitionspsychologischem, testwissenschaftlichem oder unterrichtspraktischem Blickwinkel betrachtet werden (vgl. Rossa 2012, S. 5–18). In dieser Arbeit steht Letzteres, d. h. die Sichtweise der angewandten Didaktik im Vordergrund, weil die benutzten Tests zum Hörverstehen aus Lehrwerken entnommen wurden. Der in der vorliegenden Arbeit verwendete Test zum Hörverstehen misst dementsprechend zu allen Messzeitpunkten das globale, selektive und detaillierte Hören. Dies entspricht der alltäglichen Kommunikation, in der alle drei Arten vorkommen können.

Beim globalen Hören geht es darum, eine grobe Orientierung über das Gehörte zu erhalten. Das Ziel ist die situative und thematische Einordnung des *ganzen* Texts. Der Rezipient ordnet das Gehörte einer konkreten Situation zu, erkennt folglich beispielsweise einen Dialog zwischen einem Patienten und einem behandelnden Arzt als solchen. Eine typische Aufgabe für das globale Hören besteht im Zuordnen von Dialogen zu Fotos (vgl. Barkowski 2010, S. 309; Porsch et al. 2010, S. 152; Brinitzer et al. 2013, S. 29–30; Gehring 2018, S. 96).

Beim selektiven Hören wird der Fokus vom Ganzen auf die wesentlichen Informationen gelenkt. Beim Dialog mit dem Arzt gilt es demzufolge beispielsweise herauszuhören, woran der Patient erkrankt ist oder welche Heilungsmethoden vorgeschlagen werden. Das selektive Hören umfasst auch die Erfassung der zeitlichen Reihenfolge der Geschehnisse, d. h. der Rezipient kann beurteilen, ob der Aufenthalt im Wartezimmer und das Vereinbaren eines nächsten Termins vor oder nach dem Gespräch mit dem Arzt stattfanden. Eine typische Aufgabe für das selektive Hören ist das Festlegen einer Chronologie, z. B. durch Nummerierung von Bildern, oder die Beurteilung von Informationen als richtig oder falsch.

Beim detaillierten Hörverstehen wird die volle Aufmerksamkeit der Rezipienten auf die Details gelenkt. Die hörende Person muss beispielsweise den Namen des behandelnden Arztes oder die Adresse der Praxis erkennen oder für welche Uhrzeit der nächste Termin vereinbart wird. Das detaillierte Hören kann u. a. über Multiple Choice oder offene Antworten getestet werden (vgl. Barkowski 2010, S. 309; Porsch et al. 2010, S. 152; Brinitzer et al. 2013, S. 29–30; Gehring 2018, S. 96).

4.2.4.2 Aufgabenformat

Da das Konstruieren von Tests zum Hörverstehen eine komplexe Aufgabe ist, bei der beispielsweise sprachliche Merkmale des Sprechers berücksichtigt werden müssen (vgl. Nold/Henning 2007, S. 180), wurde für diese Qualifizierungsarbeit auf bereits vorhandene Hörtexte zurückgegriffen. Sie wurden, mit freundlicher Genehmigung der jeweiligen Verlage, zwei Lehrwerken entnommen, die den Teilnehmenden nicht bekannt sein sollten: *von A bis Z – Alphabetisierungskurs / A1: Deutsch als Zweitsprache für Erwachsene* (Feldmeier 2010d) und *Deutsch in der Schweiz A1: ein Sprachkurs für Erwachsene* (Maurer 2017). Der detaillierte Hörtext wurde zu allen Messzeitpunkten wiederholt und stellt dementsprechend ein Retest-Item dar.

Der Test zum Hörverstehen wurde in den jeweiligen Kursgruppen mit allen Teilnehmenden gleichzeitig durchgeführt, wobei jeder Teilnehmer allein an einem Tisch sitzen musste. Zunächst wurde die Kursgruppe durch die Forscherin instruiert und das erste Beispiel wurde gemeinsam gelöst. Anschliessend wurde der eigentliche Test ohne Unterbrechung durchgeführt.

Mit einem ersten Hörtext wurde zu allen Messzeitpunkten das **selektive Hören** geprüft, was in Anlehnung an telc geschah. Bevor der entsprechende Hörtext abgespielt wurde, hatten die Teilnehmenden eine Minute Zeit, um den Antwortbogen mit den Bildern und den Richtig-Falsch-Sätzen zu betrachten. Die Hörtexte wurden anschliessend jeweils zweimal abgespielt, wie dies in der telc-Prüfung A1 üblich ist (vgl. telc/Goethe Institut 2017). Zudem war in der Pilotierungsphase ersichtlich geworden, dass besonders schwächere Lernende bei einmaligem Hören überfordert waren und nichts anzukreuzen vermochten, eine dreimalige Wiedergabe dagegen das Lösen der Testaufgaben zu sehr vereinfachte (s. Abschn. 5.1).

Der selektive Hörtext beim Einstufungstest, siehe Abb. 5, entstammt dem Lehrwerk *von A bis Z – Alphabetisierungskurs / A1: Deutsch als Zweitsprache für Erwachsene* (Feldmeier 2010d, S. 25). Im Unterschied zur Vorlage wurde ein zusätzliches Item erstellt (Buslinie 7) und die

Busnummer beim Bus-Item entfernt. Ausserdem wurde ein Item durch das Symbol der Post anstelle des Zeichens für Apotheke ergänzt. Diese Änderungen erfolgten aufgrund der Ergebnisse in der Erprobung, bei der sich zeigte, dass der Mehrheit der Teilnehmenden das Apotheken-Zeichen nicht bekannt war, was zu Irritationen führte und dazu, dass das Bild oft ganz ausgelassen wurde (s. Abschn. 5.1.1).


Die Texte für das selektive Hören in den beiden Follow-up-Messungen stammen aus dem Lehrwerk *Deutsch in der Schweiz A1: Ein Sprachkurs für Erwachsene* (Maurer 2017, S. 40, 91).

Wie ist die richtige Reihenfolge? Bringen Sie die Bilder in die richtige Reihenfolge (1–5).
Sie hören jeden Dialog zweimal.


Beispiel

Dialog


0




Dialog




Dialog



Dialog



Dialog



Dialog

7




Abb. 5: Selektives Hörverstehen, Einstufungstest (i. A. an Feldmeier 2010d, S. 25)

Der zweite Teil des Tests zum Hörverstehen untersuchte die Fähigkeit zum **globalen Hören**. Der entsprechende Hörtext wurde nur einmal abgespielt, erstens wiederum in Anlehnung an die A1 telc-Prüfung (vgl. telc/Goethe Institut 2017) und zweitens, weil in der Erprobung der Testteil zum globalen Hörverstehen bei zweimaliger Wiedergabe von den Teilnehmenden fast durchgehend korrekt gelöst wurde, was keine differenzierte Unterscheidung der Prüflinge mehr möglich machte (s. Abschn. 5.1).


Der globale Hörtext im Einstufungstest, siehe Abb. 6, stammt aus dem Lehrwerk *von A bis Z – Alphabetisierungskurs | A1: Deutsch als Zweitsprache für Erwachsene* (Feldmeier 2010d, S. 86).

88


Im ersten Follow-up wurde auf Hörtexte aus dem gleichen Lehrwerk zurückgegriffen (Feldmeier 2010d, S. 40, 50, 54, 62, 66, 108). Da die meisten globalen Hörtexte auf Niveau A1 eher kurz sind, d. h. max. drei Zuordnungsaufgaben beinhalten, wurden hier einige Audiodateien zusammengeschnitten. Insgesamt wurden sieben globale Hörtexte (inkl. Beispiel) ausgewählt und zu einem einzelnen Test zusammengefügt. Die Pausen zwischen den Hörtexten entsprechen den Vorgaben der telc-Prüfung A1 (vgl. telc/Goethe Institut 2017). Beim zweiten Follow-up wurde auf zwei globale Hörtexte von *Deutsch in der Schweiz A1: Ein Sprachkurs für Erwachsene* (Maurer 2017, S. 64, 75) rekurriert, die wiederum zu einem Test montiert wurden.

Welcher Dialog passt zu welchem Foto? (1–5)
Sie hören jeden Dialog einmal.


Beispiel




Dialog ☐




Dialog ☐




Dialog ☐



Dialog ☐



Dialog ☐



Dialog ☐

Abb. 6: Globales Hörverstehen, Einstufungstest (Feldmeier 2010d, S. 86)

Der dritte Hörtext war zu allen Messzeitpunkten derselbe und widmete sich dem **detaillierten Hörverstehen** (s. Abb. 7). Der entsprechende Text wurde aus dem Lehrwerk *Deutsch in der Schweiz A1: Ein Sprachkurs für Erwachsene* entnommen (Maurer 2017, S. 43) und zweimal ab Band abgespielt, in Anlehnung an telc (vgl. telc/Goethe Institut 2017) sowie um die Gegebenheiten in der Realität widerzuspiegeln, z. B. bei Ansagen am Bahnhof zu Gleisänderungen (vgl. Brinitzer et al. 2013, S. 30). Zusätzlich wurden zwei weitere Distraktoren eingebaut, damit auch hier unterschiedliche Leistungsstärken der Teilnehmenden abgebildet

werden konnten. Im ursprünglichen Hörtext des Lehrwerks ist lediglich die Hausnummer der Adresse falsch (71 statt 17). Zusätzlich wurden die Nationalität (Schweizerin statt Russin) und der Zivilstand (verheiratet statt ledig) geändert.

Kreuzen Sie an: richtig oder falsch?
Sie hören den Text zweimal.

			Anmeldung Einwohnerkontrolle Schaffhausen	
Beispiel	<input checked="" type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>Scharapowa</u>
				Name
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>Tatjana</u>
				Vorname
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>Birkenstrasse 71</u>
				Strasse, Nr.
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>8200 Schaffhausen</u>
				PLZ, Ort
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>052 624 56 26</u>
				Telefon
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>Schweizerin</u>
				Nationalität
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>29 Jahre</u>
				Alter
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>verheiratet</u>
				Zivilstand
	<input type="checkbox"/>	richtig	<input type="checkbox"/> falsch	<u>Informatikerin</u>
				Beruf
				<u>T. Scharapowa</u>
				Unterschrift

Abb. 7: Detailliertes Hörverstehen, zu allen drei Messzeitpunkten (i. A. an Maurer 2017, S. 43)

Der Test zum *Hörverstehen* dauerte in der Einstufung knappe 10 Minuten, beim ersten Follow-up 12 Minuten und beim zweiten Follow-up 18 Minuten. Dieser zeitliche Unterschied ist dadurch begründet, dass die ersten zwei Aufgaben – das selektive und das globale Hören – üblicherweise an Länge zunehmen, sobald das Niveau ansteigt.

4.2.4.3 Bewertungsverfahren

Die Bewertung des Hörverstehens erfolgte durch eine binäre Unterscheidung in richtig/falsch, wobei fehlende Angaben als Fehler gewertet wurden. Für eine korrekte Antwort gab es einen Punkt, für eine falsche oder fehlende Antwort null Punkte; halbe Punkte waren nicht möglich. Im Einstufungstest sind maximal 18 Punkte und in beiden Follow-up-Messungen maximal 19 Punkte möglich. In den beiden Follow-up-Messungen umfasst das selektive Hören, die erste Aufgabe, insgesamt sechs und nicht nur fünf Punkte.

Durch das geschlossene Aufgabenformat besteht immer die Möglichkeit, das entsprechende Feld zufällig richtig anzukreuzen. Dies führt dazu, dass das Sprachwissen eines Teilnehmenden systematisch überschätzt wird, weil für das Leistungsvermögen nur ein Teil der Punkte aussagekräftig sind und weitere Punkte durch Zufall erzielt werden. Im vorliegenden Forschungsdesign ist dies aber vernachlässigbar. Beim telc-Test verwendet das Hörverstehen ebenfalls ein geschlossenes Format, insofern ist auch dort ein gewisser Prozentsatz der korrekten Antworten auf zufälliges Ankreuzen zurückzuführen. Dies stellt die Vergleichbarkeit zwischen den hier vorgestellten Hörverstehensaufgaben und der telc-Abschlussprüfung sicher und lässt gleichzeitig auch die Vergleichbarkeit der Teilnehmenden zu, da das zufällige korrekte Antworten für alle Personen dieselbe Wahrscheinlichkeit von .50 pro Testitem aufweist.

4.2.5 Diktat

Diktate haben in der Alphabetisierungsarbeit mit erwachsenen Migranten in der Zweitsprache Deutsch mittlerweile einen festen Platz (vgl. Feldmeier 2005b, S. 12); am häufigsten Verwendung finden sie aber nach wie vor im Unterricht mit Kindern und Jugendlichen in Europa (vgl. Kazazoglu 2013, S. 1339).

Bei einem Diktat erhalten Lerner einen vorgesprochenen oder abgespielten Input in Form eines Phonems, eines Lautes, eines Wortes oder eines Satzes, den sie im Kurzzeitgedächtnis abspeichern und dann aufschreiben müssen. Das Diktat ist demnach mit dem Hörverständnis, dem Arbeitsgedächtnis und der Sprachkompetenz eng verbunden. So kann beispielsweise ein Wort nicht korrekt aufgeschrieben werden, wenn es nicht richtig gehört oder verarbeitet wurde oder es nicht in Verbindung mit dem Vorwissen gebracht werden kann (vgl. Buck 2001, S. 73; Küppers 2006, S. 91).

Englischsprachige Studien konnten zeigen, dass das Abschneiden bei Diktaten und bei geschlossenen standardisierten Sprachtests wie z. B. TOEFL (Test for English as a Foreign Language) eine deutliche Korrelation aufweist und dass Diktate die allgemeine Sprachkompetenz testen (vgl. Rahimi 2008, S. 33–34). Diese wird laut Küppers (2006, S. 90) erreicht, da die diktierten Texte aufgrund ihrer mentalen und reproduktiven Anforderungen eine gewisse Komplexität beim Niederschreiben bergen. Im Gegensatz zum Englischen existiert in der deutschen Sprache keine entsprechende Forschung, weshalb sich nicht nachweisen lässt, inwieweit diese Zusammenhänge auch für das Deutsche gelten.

Von Kritikern wird dem Diktat entgegengehalten, dass sich mit diesem Verfahren nicht der ganze aktive Schreibwortschatz des Lernenden überprüfen liesse, sondern nur ein selektiver, vom Prüfenden oder dem Forschenden vorgegebener Wortschatz (vgl. Küppers 2006, S. 91; Rahimi 2008, S. 33–34). Da in dieser Arbeit aber nicht der Wortschatz, sondern die allgemeine Sprachkompetenz geprüft werden soll, kann dieser Kritikpunkt vernachlässigt werden. Ziel des im vorliegenden Kontext benutzten Diktats ist es, zu analysieren, inwieweit (auch unbekannte) Wörter und Sätze verschriftlicht werden können.

Des Weiteren wird am Diktat die Fehlerfokussierung kritisiert (vgl. Küppers 2006, S. 91; Rahimi 2008, S. 33–34). Um dem entgegenzuwirken, wird in diesem Forschungsprojekt bei der Bewertung keine kategoriale Unterscheidung in richtig/falsch vorgenommen, sondern es werden Teilpunkte vergeben für die Laut-Buchstaben-Zuordnung und für die Verständlichkeit der niedergeschriebenen Wörter bzw. Sätze (s. Abschn. 4.2.5.3).

4.2.5.1 Testkonstrukt

Im Rahmen des Forschungsprojekts dient das Diktat dazu, die Laut-Buchstaben-Zuordnung und die Verständlichkeit von A1- und A2-Wörtern und daraus gebildeten Sätzen zu erheben. Das Laut-Buchstaben-Wissen stellt eine grundlegende Komponente für die Rechtschreibung dar und ist ihr wichtigster Prädiktor, wie Allyn und Burt (1998, S. 66) durch ihre Studien mit Erwachsenen zeigen konnten. Zudem kann durch die Bewertung der Schreibleistung beim Diktat analysiert werden, in welcher Schriftspracherwerbsphase sich die Lernenden befinden, wobei für Zweitschriftlernende insbesondere die alphabetische Phase als entscheidend gilt (vgl. Nickel 1998, S. 21; Hubertus/Nickel 2003, S. 724; Feldmeier 2010b, S. 34; Rackwitz 2016, S. 52).

4.2.5.2 Aufgabenformat

Zur Erstellung der Items wurde, wie schon bei der Elicited Imitation, auf die A1- und A2-Wörter der Häufigkeitsliste von Tschirner (2016) zurückgegriffen. Zu jedem Messzeitpunkt wurden zwei verschiedene Diktate durchgeführt: ein Wortdiktat und ein Satzdiktat. Die Satzlänge beim Satzdiktat betrug zwischen drei und fünf Wörtern; diese Angaben wurden in Anlehnung an die Canadian Language Benchmarks (2015)³² übernommen. Im Gegensatz zur Elicited Imitation

³² Das Konzept der Canadian Language Benchmarks (CLB) beinhaltet eine deskriptive Skala, die zwölf Referenzpunkte der Sprachkenntnisse in Englisch als Zweitsprache (ESL) enthält. Das Konzept befasst sich mit den Bedürfnissen und Fähigkeiten erwachsener ESL-Lerner inkl. Analphabeten.

wurden lediglich Aussagesätze diktiert, da diese Form im Kursunterricht am häufigsten vorkommt. Dabei wurde versucht, die diktierten Sätze alltagsrelevant zu gestalten.

Für den ersten Testzeitpunkt (Einstufung) wurden insgesamt 30 Wörter mit Hilfe eines Zufallsgenerators ausgewählt: 21 A1- und neun A2- Wörter. Aus je vier dieser A1- und A2- Wörter wurden analog zur Elicited Imitation Sätze gebildet, indem diese mit passenden Wörtern aus dem Korpus der 1'000 häufigsten Wörter ergänzt wurden. Dementsprechend besteht der Unterschied zwischen einem A1- und einem A2-Satz lediglich in *einem* A2-Wort, der Rest stammt aus dem A1-Wortschatz. Ein Beispiel für einen A1-Satz wäre demnach *Die (A1) Arbeit (A1) macht (A1) Spass (A1)*. Ein A2-Satz wäre beispielsweise *Ich (A1) entschuldige (A2) mich (A1)*. Diese Sätze wurden anschliessend durch eine erfahrene Deutschlehrperson darauf hin überprüft, ob sie der Lebenswelt der Zielgruppe entsprechen, und in einigen Fällen angepasst. Ebenfalls wie bei den anderen Testinstrumenten wurden im Diktat 30 % der Items beim zweiten und dritten Messzeitpunkt wiederholt, was für Wörter wie Sätze gilt. Die Auswahl der Retest-Items geschah nach dem Einstufungstest aufgrund statistischer Kriterien (s. Abschn. 6.1.2.2). Die Retest-Items sind folgende:

- D8: finden
- D9: Winter (A2-Niveau)
- D10: einfach
- D13: teilen
- D14: nochmal
- D16: stossen
- D22: Erwachsene (A2-Niveau)
- D24: Ich entschuldige mich. (A2-Niveau)
- D25: Die Arbeit macht Spass.

Im Anschluss an die Pilotierung wurden die Wörter und Sätze in einem Tonstudio abwechselnd durch eine weibliche und eine männliche Person vorgesprochen und aufgenommen. Um das Testverfahren zu standardisieren, wurden die Äusserungen zu einer einzigen Audiodatei geschnitten. Jede Tonbandaufnahme begann stets mit der Instruktion, gefolgt von einem Beispiel, das auf dem Testdokument ersichtlich war, und einem Beispiel, das durch die Prüflinge selbst verschriftlicht werden musste. Diese Verschriftlichung wurde vor dem eigentlichen Testbeginn bei jedem Lernenden durch die Forscherin überprüft, um sicherzustellen, dass alle Teilnehmenden das Testverfahren verstanden hatten. Im Anschluss daran begann das Diktat. Jedes Wort wurde insgesamt viermal diktiert mit einer Pause von 3 Sekunden zwischen jeder Wiederholung. Das nächste Wort wurde jeweils von einem Tonsignal angekündigt. Die Sätze wurden mit gleichem zeitlichen Abstand diktiert, wobei jeder Satz fünfmal wiederholt wurde. Obwohl die Items im Diktat üblicherweise lediglich ein- bis

dreimal diktiert werden (vgl. Scheibner-Herzig et al. 1991, S. 230; Buck 2001, S. 73–74; Rackwitz 2016, S. 90), wurde der Input in diesem Test vier- bzw. fünfmal wiederholt. Dieses Vorgehen wurde gewählt, damit das Arbeitsgedächtnis weniger belastet wird; die gehörten Informationen mussten dadurch nicht so lange abgespeichert werden.

Um das Testverfahren für die Lernenden so einfach wie möglich zu halten, wurde den Lernenden eine Vorlage für das Niederschreiben der Wörter und Sätze ausgehändigt (s. Abb. 8).

Vorname: _____

Familienname: _____

Hören Sie und schreiben Sie die Wörter. Sie hören die Wörter viermal.

Beispiel
1. Ende
2. _____

1. _____

2. _____

3. _____

Abb. 8: Diktatvorlage (Ausschnitt)

Das Diktat im Einstufungstest beinhaltete insgesamt 22 Wörter; fünf A2-Wörter und 17 A1-Wörter, sowie acht Sätze, jeweils vier im A2- und A1-Bereich (s. Tab. 21). Das Diktat im Einstufungstest dauerte insgesamt 13 Minuten.

	Wort	Silbenzahl	Niveau
1.	Zeit	1	A1
2.	Welt	1	A1
3.	Haus	1	A1
4.	schlecht	1	A1
5.	Loch	1	A2
6.	fallen	2	A1
7.	genug	2	A1
8.	finden	2	A1
9.	Winter	2	A2
10.	einfach	2	A1
11.	davon	2	A1
12.	Produkt	2	A1
13.	teilen	2	A1
14.	nochmal	2	A1
15.	vorbei	2	A2
16.	stossen	2	A1
17.	tragen	2	A1
18.	lieber	2	A1
19.	Artikel	3	A1
20.	einschränken	3	A2
21.	ermöglichen	4	A1
22.	Erwachsene	4	A2
	Äusserung	Wortzahl	Niveau
1.	Ich melde mich.	3	A1
2.	Ich entschuldige mich.	3	A2
3.	Die Arbeit macht Spass.	4	A1
4.	Sein Lehrer ist toll.	4	A1
5.	Das Gebäude ist alt.	4	A2
6.	Er hat kleine Finger .	4	A2
7.	Ich erkläre dir das Wort.	5	A1
8.	Sie muss die Aufgabe verbessern .	5	A2

Tab. 21: Äusserungen Diktat, Einstufung
Kursive Ziffer: Retest-Item; fettgedruckt: A2-Wort

Das Diktat im Follow-up 1 nach 101 Kurslektionen beinhaltete neun Retest-Items, die aus der Einstufung stammten, sieben Wörter und zwei Sätze. Dazu kamen insgesamt 15 neue Wörter, drei A2-Wörter und zwölf A1-Wörter, sowie sechs neue Sätze, je drei im A2- und im A1-Bereich (s. Tab. 22). Das Diktat im Follow-up 1 dauerte ebenfalls knappe 13 Minuten.

	Wort	Silben	Niveau
1.	ihr	1	A1
2.	Film	1	A1
3.	Mund	1	A1
4.	falsch	1	A1
5.	Start	1	A2
6.	finden	2	A1
7.	heute	2	A1
8.	einfach	2	A1
9.	Mutter	2	A1
10.	nochmal	2	A1
11.	System	2	A1
12.	stossen	2	A1
13.	wachsen	2	A1
14.	Kunde	2	A1
15.	teilen	2	A1
16.	offen	2	A1
17.	Winter	2	A2
18.	Beitrag	2	A1
19.	Polizei	3	A1
20.	aufhören	3	A2
21.	kontrollieren	4	A2
22.	Erwachsene	4	A2
	Äusserung	Wortzahl	Niveau
1.	Es ist heiss.	3	A1
2.	Ich vertraue dir.	3	A1
3.	Ich entschuldige mich.	3	A2
4.	Sie hat sich verletzt .	4	A2
5.	Er hat auch kalt.	4	A1
6.	Die Arbeit macht Spass.	4	A1
7.	Ich gehe die Zeitung holen.	5	A1
8.	Ich suche die richtige Definition .	5	A2

Tab. 22: Äusserungen Diktat, Follow-up 1
Kursive Ziffer: Retest-Item; fettgedruckt: A2-Wort

Das Diktat im Follow-up 2 nach 200 Kurslektionen umfasste nebst den Retest-Items 15 neue Wörter, drei A2-Wörter und zwölf A1-Wörter, sowie wiederum sechs neue Sätze (s. Tab. 23). Auch das Diktat im Follow-up 2 dauerte insgesamt knappe 13 Minuten.

	Wort	Silben	Niveau
1.	Raum	1	A1
2.	vorn	1	A1
3.	klein	1	A1
4.	finden	2	A1
5.	können	2	A1
6.	einfach	2	A1
7.	Doktor	2	A1
8.	nochmal	2	A1
9.	Technik	2	A1
10.	stossen	2	A1
11.	feiern	2	A2
12.	sprechen	2	A1
13.	Winter	2	A2
14.	leben	2	A1
15.	teilen	2	A1
16.	Sommer	2	A1
17.	beide	2	A1
18.	vorhanden	3	A1
19.	erlauben	3	A1
20.	erschrecken	3	A2
21.	verheiratet	4	A2
22.	Erwachsene	4	A2
	Äusserung	Wortzahl	Niveau
1.	Ich entschuldige mich.	3	A2
2.	Die Klasse ist ruhig.	4	A1
3.	Ich unternehme eine Reise.	4	A2
4.	Die Arbeit macht Spass.	4	A1
5.	Sie hat einen neuen Mitarbeiter.	5	A1
6.	Sein Sohn ist kein Arzt.	5	A1
7.	Die rote Tasche passt gut.	5	A2
8.	Der weisse Boden ist kalt.	5	A1

Tab. 23: Äusserungen Diktat, Follow-up 2
Kursive Ziffer: Retest-Item; fettgedruckt: A2-Wort

4.2.5.3 Bewertungsverfahren

Für die Bewertung des Diktats wurde das Bewertungsraster von Backhaus und Rackwitz (2011a, S. 35) übernommen (s. Tab. 24), zusätzlich aber auch die Verständlichkeit der Schreibweise bewertet.

Die Bewertung erfolgte einerseits durch eine erfahrene Deutschlehrperson und andererseits durch die Forscherin selbst. Bei Unstimmigkeiten wurde stets eine dritte Person hinzugezogen. Lernende konnten maximal fünf Punkte für die Laut-Buchstaben-Zuordnung und maximal zwei Punkte für die Verständlichkeit erreichen. Somit konnten pro Testzeitpunkt maximal 150 Punkte in der Laut-Buchstaben-Zuordnung und maximal 60 Punkte in der Verständlichkeit erzielt werden, was einem Total von 210 Punkten entspricht. Bei der Interpretation der Resultate wurden die Werte für Verständlichkeit und Laut-Buchstaben-Zuordnung auf 100 %

normiert, d. h. beide Elemente werden bei der Auswertung der Resultate als gleichwertig betrachtet.

Punkte	Verständlichkeit	Laut-Buchstaben-Zuordnung	Anknüpfend an das Modell von Günther
5		Das Wort wird orthografisch korrekt geschrieben.	Orthografische Phase
4		Die Lautfolge wird komplett abgebildet.	Orthografische Phase
3		Die Lautfolge wird fast komplett (mehr als 2/3) abgebildet.	Alphabetische Phase
2	leicht verständlich, ohne Interpretation.	Das ‚Lautskelett‘ (weniger als 2/3 der Laute zutreffend) wird abgebildet.	Alphabetische Phase
1	verständlich nur durch Interpretation.	Ein Laut (meist Anlaut) wird durch einen passenden Buchstaben wiedergegeben.	Logografische Phase
0	völlig unverständlich.	Kein Laut eines Buchstabens wird durch einen passenden Buchstaben wiedergegeben (willkürliche Schreibweise).	Logografische Phase

Tab. 24: Bewertungsraster Diktat (i. A. an Backhaus/Rackwitz 2011b, S. 35–36)

Zum besseren Verständnis soll die Bewertung anhand eines Beispiels erläutert werden.

Die Teilnehmenden hörten jeweils einen Stimulus den sie auf die Vorlage schreiben sollten, beispielsweise:

‚Winter‘ (Retest-Item: ED9, F1D17 und F2D13)

Die korrekte Verschriftlichung wurde mit der vollen Punktzahl bewertet. Die Variante ‚Wintr‘ wurde mit vier von fünf Punkten in der Laut-Buchstaben-Zuordnung (L-B-Z) und zwei von zwei Punkten in der Verständlichkeit bewertet, da die Lautfolge zwar korrekt abgebildet ist, aber die Buchstabenfolge nicht korrekt ist. ‚Wenter‘ erhielt drei Punkte in der L-B-Z und zwei Punkte in der Verständlichkeit, weil die Lautfolge mit einer Ausnahme fast komplett abgebildet ist. Für ‚mentr‘ wurden zwei Punkte in der L-B-Z und null Punkte in der Verständlichkeit vergeben, da weniger als zwei Drittel der Laute korrekt verschriftlicht wurden und das Wort nicht verständlich ist. ‚Atl‘ wurde mit einem Punkt in der L-B-Z und null Punkten in der Verständlichkeit bewertet, da lediglich ein Laut durch einen korrekten Buchstaben verschriftlicht wurde und das Wort hier ebenfalls nicht verständlich ist. Null Punkte gab es, wenn nichts geschrieben wurde oder bei willkürlichen Schreibweisen wie ‚boF‘.

4.3 Fragebögen

4.3.1 Lehrerfragebogen

Um Informationen über die Lehrperson, ihren beruflichen Werdegang, ihre Erfahrungen im Unterrichten sowie ihre Meinung über das Kurskonzept zu erfahren, wurde den Lehrpersonen vor Kursbeginn ein Fragebogen vorgelegt (s. Anhang, Kap. 7, S. 225). Dieser umfasste 19 Fragen zur Person (Geschlecht, Alter, Nationalität, Erstsprache und

Fremdsprachenkenntnisse) und der bisherigen Lehrtätigkeit (Lehrerfahrung DaZ und Alphabetisierung, Hochschulabschlüsse und Weiterbildungen). Anhand dieser personenbezogenen Fragen sollte eine Einschätzung der Kursleitenden und ihres Bildungs- und Erfahrungshintergrunds vorgenommen werden, um diese Angaben auf mögliche als ‚Ausreisser‘ einzustufende Merkmale hin kontrollieren zu können. Zusätzlich wurden die Lehrpersonen gebeten, schriftlich festzuhalten, ob sie die im *Zweitschriftlernerkonzept* anvisierten Ziele für realistisch halten. Das Ausfüllen des Fragebogens dauerte fünf bis zehn Minuten.

4.3.2 Lernerfragebogen

Der Fragebogen für Lernende (s. Anhang, Kap. 8, S. 231) umfasst in allen Sprachen 17 Fragen zum persönlichen Hintergrund der Teilnehmenden (Alter, Geschlecht, Erstsprache, Kinder, Gesundheitszustand), zur Schul- und Arbeitsbiografie (Anzahl Schuljahre, höchster Schulabschluss, Ausbildungsabschluss, Arbeitserfahrung) und zu den Sprachkenntnissen (besuchte Deutschkurse, L1-Schreib- und Lesekenntnisse, Fremdsprachen, L2-Lese- und Schreibkenntnisse, Gebrauch des Deutschen). Anhand dieser Fragen konnten verschiedene personenbezogene Informationen über die Teilnehmenden erfasst werden, in Anlehnung an die Merkmale für Zweitschriftlernende gemäss BAMF (vgl. BAMF 2018a, S. 10–11).

Nebst diesen Fragen wurden die Teilnehmenden aufgefordert, einen kurzen Text in ihrer Erstschriftsprache über sich oder ihre Familie zu verfassen, wobei die Angaben fiktiv sein durften. Dieser Teil diente der Überprüfung der Schreibkompetenz in der Erstsprache bzw. der Erstschriftsprache der Lernenden, die durch versierte Übersetzer anhand eines Bewertungsrasters beurteilt wurde.

Bei der Bewertung von Lernertexten in deren Erstschriftsprache wird häufig auf Übersetzer verzichtet und die Bewertung wird durch die Lehrperson bzw. die zuständige Einstufungsperson selbst vorgenommen. Anstelle der tatsächlichen L1-Kenntnisse werden vielmehr Merkmale wie Fehlhaltung des Stifts oder Frustration und Schwierigkeiten beim Niederschreiben festgehalten und anschliessend durch die Einstufungsperson interpretiert (vgl. Florida Department of Education 2014, S. 2). Diese Merkmale können jedoch sowohl bei Analphabeten wie auch bei Zweitschriftlernenden auftreten. So muss beispielsweise ein falsches Halten des Schreibgeräts nicht zwingend auf ungenügende Schreibkompetenzen zurückzuführen sein, da in einigen Sprachen der Stift auch bei korrekter Schreibweise beim

Schreiben eines Buchstabens mehrfach abgesetzt werden muss (vgl. Guerrero Calle im Erscheinen).

Die Bewertung der Schreibkompetenz anhand der letzten Aufgabe im Fragebogen wurde in dieser Studie daher erfahrenen Übersetzern überlassen. Diese wurden anhand dreier Kriterien ausgewählt: Sie sollten selbst aus dem jeweiligen Herkunftsland stammen, damit länderbezogene Dialekte berücksichtigt werden konnten, und sie sollten ein hohes Deutschniveau (mind. C1) sowie selbst einen Fluchthintergrund aufweisen, damit sie eine entsprechende Sensibilisierung mitbrachten. Praktisch alle Übersetzer mit Ausnahme des Persischen besaßen zudem ein Schweizer Zertifikat als interkulturelle Dolmetscher. Damit die Bewertung aller Texte einheitlich erfolgte, haben die Übersetzer alle mit demselben Bewertungsraster und mit der Forscherin zusammengearbeitet.

Wie in Tab. 25 dargestellt, konnten die Lernenden bei der Beurteilung ihrer Schreibkompetenz in der L1 null bis drei Punkte erreichen. Nur Personen, die mindestens einen Punkt erreichten, d. h. mindestens Niveau A2, wurden zur Studie zugelassen. Zwei Punkte bedeuten, dass der Übersetzer ein Niveau auf B1 und darüber attestierte und drei Punkte implizieren ein Niveau von C1 und darüber.

Punkte/ Niveau nach GER	Wortschatzbeherrschung	Formale Richtigkeit (Grammatik und Orthografie)
3 C1/C2	Gelegentlich kleinere Schnitzer, aber keine grösseren Fehler im Wortgebrauch.	[...] hohes Mass an grammatischer Korrektheit[...]. Die Rechtschreibung ist, abgesehen von gelegentlichen Verschreibern, korrekt.
2 B1/B2	[...] gute Beherrschung des Grundwortschatzes, macht aber noch elementare Fehler [...].	[...] Gute Beherrschung der grammatischen Strukturen [...]. Rechtschreibung, Zeichensetzung und Gestaltung sind exakt genug, sodass man sie meistens verstehen kann.
1 A2	Beherrscht einen begrenzten Wortschatz [...].	[...] macht [...] elementare [Grammatik]fehler [...]. Kann kurze Wörter aus seinem mündlichen Wortschatz 'phonetisch' einigermaßen akkurat schriftlich wiedergeben [...].

Tab. 25: Bewertungsraster Lernerfragebogen (Trim et al. 2009, S. 113–114, 118)

Als Mindestniveau zur Kurszulassung wurde A2 festgelegt, da dies gemäss Konzept für einen bundesweiten Alphabetisierungskurs dem Niveau des funktionalen Analphabetismus nahekommen soll (vgl. Feldmeier 2015, S. 48). Die Festlegung ist jedoch nicht unproblematisch, denn, wie bereits in Abschnitt 1.1.3 besprochen, ist funktionaler Analphabetismus immer auch vom Herkunftsland abhängig, was der Idee einer fixen Einteilung für alle Länder widerspricht. Dennoch wurde diese Vorgabe hier als Referenzniveau verwendet, da weder auf andere Referenzwerte noch auf länderspezifische Werte zurückgegriffen werden kann.

In einigen Fällen war es trotz Bewertungsraster schwierig, eine adäquate Einschätzung vorzunehmen, da manche Teilnehmende sehr wenig geschrieben haben. In diesen Fällen mussten die Übersetzer entscheiden: Wenn die Person ihrer Einschätzung nach im Alltag des jeweiligen Landes ohne Schwierigkeiten mit ihren Schriftkompetenzen zurechtkommen würde, wie dies auch im Konzept für Zweitschriftlernende verlangt wird (vgl. BAMF 2018a, S. 10), wurde sie zum Kurs zugelassen.

Durch die Vorgabe des Mindestniveaus A2 konnten sieben Teilnehmende, die am Einstufungstest teilnahmen, den Zweitschriftlernerkurs nicht besuchen. Zusätzlich wurden sechs Teilnehmende abgelehnt, da diese in einem lateinischen Schriftsystem (Somali, Saho, Bilen) alphabetisiert waren und somit nicht als Zweitschriftlernende gelten.

5. Datengrundlage

Wie die Daten erhoben, kodiert und aufbereitet wurden, wird in diesem Kapitel dargestellt. Nach Besprechung der Pilotierung der fünf Tests (s. Abschn. 5.1) wird auf den genauen Testablauf eingegangen (s. Abschn. 5.2). Danach wird die Kodierung der Variablen besprochen, d. h. die abhängigen und unabhängigen Variablen werden im Detail vorgestellt (s. Abschn. 5.3). Die Datenaufbereitung und Bereinigung der Datengrundlage bilden den Abschluss des Kapitels (s. Abschn. 5.4).

5.1 Pilotierung und Überarbeitung

Bevor das Testverfahren für die Teilnehmenden der Zweitschriftlernerurse angewendet werden konnte, wurde es erprobt. Geprüft wurden einerseits die Verständlichkeit der Testanlagen und die Klarheit der Instruktionen. Andererseits wurden alle Items erprobt, um diese anschliessend mittels einer klassischen Itemanalyse statistisch auszuwerten und gegebenenfalls anzupassen.

Die Pilotierung der fünf Testinstrumente erfolgte zwischen April und Mai 2018 in regulären Alphabetisierungskursen. Das Testverfahren wurde jeweils mit allen anwesenden Teilnehmenden durchgeführt, unabhängig davon, ob sie primäre Analphabeten oder Zweitschriftlernende waren. Dieses Vorgehen hat ethische Gründe, um den primären Analphabeten im Kurs nicht das Gefühl zu vermitteln, dass sie wegen mangelnder Lese- und Schriftkenntnisse in ihrer Erstschriftsprache diskriminiert würden. Tatsächlich ausgewertet wurden jedoch nur die Ergebnisse der Zweitschriftlernenden. An dieser Stelle ist anzumerken, dass primäre Analphabeten deutlich mehr Instruktion benötigten als Zweitschriftlernende und ihr Arbeitstempo entscheidend langsamer war. Dies ist ein weiterer Hinweis darauf, primäre Analphabeten und Zweitschriftlernende separat zu unterrichten.

Es war ursprünglich geplant, die Tests mit einer grösseren Stichprobe zu pilotieren, um dadurch die Testgüte zu steigern. Wie sich jedoch herausstellte, war die Anzahl der verfügbaren Zweitschriftlernenden geringer als erwartet, weshalb die Pilotierung mit weniger Personen durchgeführt werden musste als vorgesehen. Im Folgenden ist deshalb von Erprobung statt Pilotierung die Rede.

5.1.1 Erste Erprobung

In der ersten Erprobung konnten Ergebnisse von 16 Zweitschriftlernenden berücksichtigt werden: von je fünf Zweitschriftlernenden für den Einstufungstest und das erste Follow-up-

und von sechs Zweitschriftlernenden für den zweiten Follow-up-Test. Es stellte sich heraus, dass das Testverfahren allen Personen verständlich war, weshalb keine Änderungen am Testprozess und an den Instruktionen vorzunehmen waren. Anhand der mit SPSS berechneten Trennschärfe wurde entschieden, welche Items des **Diktats** verändert werden mussten. Da die Stichprobe der Erprobung sehr klein war (fünf oder sechs Personen je Item) und dadurch eine Itemanalyse nur bedingt zuverlässig ist, wurde festgelegt, dass nur Items mit negativer Trennschärfe und somit ganz klare Ausreisser ausgetauscht werden. Vor jeder Änderung wurden die Items mit erfahrenen Lehrpersonen besprochen, denn, wie Bühner (2011, S. 81) schreibt, ist das Ausscheidkriterium auch jeweils stark vom Kontext abhängig. Folgende Items wurden ausgetauscht (s. Tab. 26):

Messzeitpunkt		Ursprüngliches Item	Trennschärfe	Schwierigkeits-index	Neues Item
Einstufungstest <i>n</i> = 5	Diktat	gegen	-.16	.16	einfach
		Erinnerung	-.54	.04	ermöglichen
		Die Situation ist schwierig.	-.36	.16	Sein Lehrer ist toll.
Follow-up 1 <i>n</i> = 5	Diktat	falsch	-.48	.60	erleben
		Erklär mir das Wort.	-.62	.44	Er hat auch kalt.
		Die Prüfung ist schwierig.	-.14	.40	Sie hat sich verletzt.
		Ich möchte meine Meinung sagen.	-.57	.44	Ich gehe die Zeitung holen.
Follow-up 2 <i>n</i> = 6	Diktat	an	-.16	.30	vorn
		liegen	-.41	.46	klein
		Haus	-.17	.96	sprechen
		auch	-.96	.94	können
		wiederum	-.41	.46	beide

Tab. 26: Itemanalyse Diktat, Erprobung 1

Bei der Erprobung des **Hörverstehens** wurde ersichtlich, dass das selektive Hören zweimal durchgeführt werden sollte und dafür das globale Hören lediglich einmal. Schwächere Teilnehmende konnten im selektiven Teil bei einmaligem Hören praktisch nichts ankreuzen und bei dreimaliger Wiedergabe des Textes fiel der Test eher zu einfach aus. Im globalen Hören war bei zweimaligem Abspielen des Tonträgers kaum noch eine Unterscheidung der Teilnehmenden möglich, da fast jeder alles korrekt zu lösen vermochte. Zusätzlich musste im Bereich des selektiven Hörens bei der Einstufung das Bild des Apothekenzeichens durch das Postzeichen ersetzt werden. Vier der fünf Teilnehmenden haben dieses Bild bei der Beantwortung ausgelassen. Möglicherweise hat der Umstand, dass der Protagonist des Hörtexts bei der Post und nicht bei der Apotheke aussteigt, zu Verwirrung geführt.

Wahrscheinlicher ist aber, dass den Prüflingen das deutsche und in der Schweiz nicht gebräuchliche Apothekenzeichen unbekannt war.

Der dritte Teil des Tests zum Hörverstehen (detailliert; Formular korrigieren) wurde in allen Stufen verstanden. Obwohl dieser Teil für die Einstufung etwas schwierig war, wurde er nicht geändert, da er ebenfalls als Retest-Item in allen Messzeitpunkten diente.

Die **Kunstwörter** wurden insgesamt mit vier Zweitschriftlernenden erprobt. Bei der Erprobung zeigte sich, dass es die Teilnehmenden verunsicherte, wenn lediglich das falsch nachgesprochene Kunstwort durch die Testleiterin aufgeschrieben wurde. Dementsprechend wurden in den späteren Testsituationen alle wiederholten Kunstwörter aufgeschrieben und nicht nur die falsch wiederholten. Zusätzlich wurden die Kunstwörter von acht weiteren Personen ab Tonband angehört und niedergeschrieben, darunter vier Zweitschriftlernende mit guten Kenntnissen in der deutschen Sprache und vier Personen mit deutscher Erstsprache. Damit sollte überprüft werden, ob die Tonaufnahme von hinreichender Qualität ist bzw. ob die Aussprache der aufgenommenen Wörter deutlich genug ist. Es hat sich ergeben, dass folgende Silben (hier fettgedruckt) der Kunstwörter ***dugabe***, ***relomano***, ***gebidadfino***, ***ronamelita*** und ***bigadonafera*** nur ungefähr von der Hälfte des zusätzlichen Erprobungssamples richtig reproduziert werden konnten. Diese Silben wurden erneut aufgenommen und in der Audioaufnahme ausgetauscht.

Die Items der **Elicited Imitation** wurden von allen Probanden verstanden und führten nicht zu weiteren Schwierigkeiten. Hier wurde auf eine Itemanalyse verzichtet, da anzunehmen war, dass die Probanden über höhere mündliche Sprachkenntnisse verfügten als die eigentliche Zielgruppe.

Die Texte der **ORF-Tests** konnten ebenfalls problemlos durch die Teilnehmenden vorgelesen werden und machten keine weitere Anpassung notwendig.

Bei der Erprobung der **Fragebögen** hat sich gezeigt, dass auf Tigrinya die Fragen zu Betreuungspflichtigen Kindern, Schulabschluss und Gebrauch des Deutschen ausserhalb des Unterrichts zu Missverständnissen führten. Zusätzlich hatten die Testpersonen ihren Namen in ihrem Schriftsystem geschrieben, sodass er für die Forscherin nicht lesbar war. Das Ausfüllen des Fragebogens nahm jeweils 15 Minuten in Anspruch. Die genannten Fragen wurden daraufhin erneut mit dem Übersetzer besprochen und in einfacherer Sprache verfasst. Zudem wurden die Probanden bei den effektiven Testdurchführungen darauf hingewiesen, den Namen in lateinischen Buchstaben zu schreiben.

Das Zeitmanagement musste nicht angepasst werden, da Zweitschriftlernende alle Aufgaben innerhalb der geplanten Zeit lösen konnten.

5.1.2 Zweite Erprobung

Nach den Anpassungen der ersten Erprobung wurden die geänderten Items des Diktats, des Hörverstehens und der Kunstwörter mit anderen Kursteilnehmenden erprobt. Auch hier waren die Kurse durchmischt, d. h. es gab primäre Analphabeten und Zweitschriftlernende im jeweiligen Kurs. Analysiert wurden lediglich die Antworten der Zweitschriftlernenden: von fünf Zweitschriftlernenden für den Einstufungstest, von sechs Zweitschriftlernenden für den ersten Follow-up- und von sieben Zweitschriftlernenden für den zweiten Follow-up-Test. Auf Wunsch der Kursleitenden wurde diesmal nicht das ganze Diktat im Unterricht erprobt, sondern nur die aufgrund der ersten Erprobung geänderten Items. Aus diesem Grund und weil die Stichprobe der zweiten Erprobung erneut relativ klein war, wurde auf ein prüfstatisches Verfahren verzichtet. Die Resultate der zweiten Erprobung wurden aber auf offensichtliche Ausreisser kontrolliert. Da keine solche Ausreisser entdeckt wurden, wurden keine weiteren Anpassungen an der Itemauswahl vorgenommen.

Beim **Hörverstehen** hat sich die Änderung des Apothekenzeichens in ein Postzeichen als sinnvoll erwiesen, d. h. die Aufgabe konnte nun komplett gelöst werden. Zudem erwies sich das einmalige Abspielen des globalen Hörtexts und eine zweimalige Wiedergabe des selektiven Texts bezüglich der Schwierigkeit als gut praktikabel.

Die neu aufgenommenen **Kunstwörter** wurden bei der zweiten Erprobung deutlich besser reproduziert. Einige Personen hatten nach wie vor Mühe, wofür besonders die Länge der Kunstwörter ein Grund sein könnte.

5.2 Testablauf

Nach Darlegung der jeweiligen Testkonstruktion und der Erprobung wird im Folgenden auf das Testverfahren in der Endversion eingegangen.

Das vorstehend detailliert beschriebene Testverfahren soll in dieser Forschungsarbeit potenzielle Unterschiede in Teilbereichen der Sprachkompetenz – Sprechen, Leseflüssigkeit, Hörverstehen und Laut-Buchstaben-Zuordnung – zwischen den Treatmentgruppen sichtbar machen. Jedes einzelne Testinstrument überprüft demnach ein bestimmtes Phänomen wie z. B. die Leseflüssigkeit anhand der ORF. Dazu wurde das Testverfahren jeweils vor Kursbeginn, dann nochmals nach 100 und nach 200 Kurslektionen eingesetzt.

Das Testverfahren umfasst zwei Teile: einen mündlichen, individuellen Teil 1, bestehend aus Elicited Imitation, Kunstwörter und Oral Reading Fluency, und einen schriftlichen, gemeinsamen Teil 2, bestehend aus Hörverstehen und Diktat. Die Tests fanden jeweils während der regulären Kurszeiten der Teilnehmenden statt.

Für den ersten Teil wurden immer drei zu Prüfende aus dem Unterricht abgeholt, um individuell in drei separaten Räumen jeweils von einem Prüfungsleiter auf ihre mündlichen Kompetenzen hin getestet zu werden. Nach der Überprüfung des Elicited Imitation, der Kunstwörter und der Oral Reading Fluency wurden die Testpersonen ins Kurszimmer zurückgeführt und drei neue Personen abgeholt. Für alle übrigen Lernenden wurde der Unterricht wie gewohnt fortgesetzt. Pro Person dauerte dieser Teil des Tests ungefähr 20 Minuten. Der mündliche Teil wurde mit Hilfe eines digitalen Audioaufnahmege­räts (Olympus WS-853) aufgenommen und für die spätere Auswertung abgespeichert.

Nachdem Teil 1 für alle abgeschlossen war, wurde der reguläre Unterricht pausiert, damit alle Prüflinge gemeinsam den zweiten, schriftlichen Teil lösen konnten. Dies geschah unter Supervision der Forscherin und dauerte ungefähr 30 Minuten.

Für den ersten Teil des Tests wurden mehrere Testleiter benötigt, um die individuelle Testdurchführung mit den Teilnehmenden parallel und damit zeiteffizient zu gestalten. Die Testleiter hatten vorgängig mündliche Instruktionen und ein Dokument zur Prüfungsdurchführung erhalten. Als unentgeltliche Testleiter gewonnen werden konnten einerseits der Forscherin bekannte Personen, die im Rahmen ihrer professionellen Tätigkeit mit der Zielgruppe in Kontakt standen. Andererseits konnten Studierende des Bachelors Deutsch als Fremdsprache an der Universität Fribourg als Testleiter eingesetzt werden. Im Gegenzug für ihre Mitarbeit am Forschungsprojekt durften sie eine verkürzte Prüfung in einem Seminar bei Prof. Dr. Thomas Studer ablegen.

Der Einstufungstest fand bei beiden Institutionen an verschiedenen Tagen im Juli 2018 statt. Die Follow-up-Messungen erfolgten ganztägig bei Institution P am 05.09.2018 und am 31.10.2018 und bei Institution Q am 18.09.2018 und am 06.11.2018.

Tab. 27 fasst das Untersuchungsdesign zusammen: Die Kurseinteilung wurde im Anschluss an den Einstufungstest vorgenommen, anschliessend wurde nach 101 und nach 200 Kurslektionen je eine Follow-up-Messung durchgeführt und der Kurs wurde mit der offiziellen telc-A1-Prüfung nach 300 Lektionen beendet.

Untersuchungsdesign Versuchsplan							
Zwei Vergleichsgruppen mit Prätest, 2 Follow-ups und Posttest							
Juni	Juli	August	September	Oktober	November	Dezember	
O1	K	X1	O2	X1	O3	X1	O4
O1	K	X1, 2	O2	X1, 2	O3	X1, 2	O4
K: Kurseinteilung, X: durchgehendes Treatment, O: Messung der AV/Testung							

Tab. 27: Versuchsplan (i. A. an Darsow/Felbrich 2014, S. 235)

5.3 Kodierung der Variablen

5.3.1 Abhängige Variablen

In der vorliegenden Arbeit fungieren zwei Variablen als abhängige Variablen. Diese zu erklärenden Variablen sind einerseits das *telc-Resultat* und andererseits der *Gesamtlernerfolg* im Kursverlauf.

Die Testteilnehmenden absolvierten während der Studie insgesamt vier Tests: einen Einstufungstest zu Beginn, je einen Follow-up-Test nach 101 respektive 200 Kurslektionen und den *telc-Abschlusstest* nach 300 Lektionen. Zur Berechnung der Höhe der Variable **telc-Resultat** wird das Testresultat der Teilnehmenden beim letzten Test, der A1-Prüfung von *telc*, herangezogen. Die *telc-Prüfung* setzt sich aus vier Teilen zusammen, Hören, Lesen, Schreiben und Sprechen. Bei allen Teilen lassen sich jeweils 25 Punkte erreichen, insgesamt sind 100 Punkte in der A1-Prüfung erreichbar, wobei die Prüfung ab 60 Punkten als bestanden gilt. Die Variable *telc-Resultat* bewegt sich somit immer zwischen 0 und 100.

Die zweite abhängige Variable, der **Gesamtlernerfolg**, gibt die Lernprogression der Teilnehmenden zwischen Beginn der Studie und dem zweiten Messzeitpunkt nach 200 Kurslektionen an. Sie wird berechnet als [Punktzahl Follow-up 2] minus [Punktzahl Einstufungstest]. Die Punktzahl der jeweiligen Testzeitpunkte wird berechnet als ein auf 100 normiertes Mittel der vier zu gleichen Teilen gewichteten Tests Elicited Imitation, Oral Reading Fluency, Hörverstehen und Diktat. Die Kunstwörter werden nicht berücksichtigt, da diese das Arbeitsgedächtnis messen und nicht die allgemeine Sprachkompetenz. Da die Probanden im Prinzip zum ersten Prüfzeitpunkt mehr Punkte erreichen können als zum zweiten, sind auch negative Werte möglich. Die metrische Variable *Gesamtlernerfolg* kann somit einen Wert zwischen minus 100 und 100 annehmen.

5.3.2 Unabhängige Variablen

Die erklärenden Variablen wurden literaturbezogen und aufgrund theoretischer Überlegungen und praktischer Erfahrungswerte gebildet und in Anlehnung an das Faktorenkomplex-Modell von Edmondson (1984) gruppiert, das soziopolitische,

wissenschaftliche, personenbezogene, unterrichtsbezogene und lehr- und lernumgebungsbezogene Faktoren umfasst. Wissenschaftliche Faktoren werden nicht als gesonderte Variablen ausgewiesen, da wissenschaftliche Erkenntnisse bereits in die Bildung der übrigen Variablen eingeflossen sind. Beispielsweise ist aus Studien bekannt, dass die Anzahl an Schuljahren einen Einfluss auf das Sprachenlernen hat (s. Abschn. 1.3.2), diese Variable wurde aber nicht unter den wissenschaftlichen Faktoren, sondern unter den personenbezogenen Faktoren aufgeführt. Soziopolitische Variablen wurden nicht gebildet. Insgesamt lassen sich drei Variablengruppen und 18 Variablen unterscheiden (s. Abb. 9). Die Kursteilnehmenden wurden durch die Forscherin den Kursgruppen zugeteilt (s. Abschn. 3.3.1), d. h., die Variablen *Institution*, *Kurslevel* und *Silbenmethode* sind durch diese Kurszuteilung vorgegeben. Die restlichen erklärenden Variablen wurden, mit Ausnahme der *Kursabsenz*, mittels Fragebogen ermittelt. Details zur Kodierung von Dummy-Variablen, und wie die Referenzkategorien zu verstehen sind, sind dem Abschnitt 5.3.3 zu entnehmen.

Lehr- und Lernumgebungsvariablen	Unterrichtsbezogene Variablen	Personenbezogene Variablen
<ul style="list-style-type: none"> •Institution 	<ul style="list-style-type: none"> •Kurslevel •Silbenmethode •Besuchte Deutschkurse 	<ul style="list-style-type: none"> •Geschlecht •Alter •L1 •Wortschatzbeherrschung Erstschrift •Formale Richtigkeit Erstschrift •Fremdsprache •Schuljahre •Berufslehre •Universität •Arbeit Heimatland •Kinder •Gesundheitszustand •Deutschgebrauch •Kursabsenz

Abb. 9: Gruppierte unabhängige Variablen (i. A. an Edmondson/House 2011, S. 25)

5.3.2.1 Lehr- und Lernumgebungsvariablen

In welcher Institution die Teilnehmenden im Rahmen der Studie unterrichtet wurden, wird durch die Variable *Institution* abgebildet (s. Tab. 28). Die Zuteilung der Teilnehmenden erfolgte auf Grundlage der in Abschnitt 3.3.1 dargestellten Kriterien durch die Forscherin.

Variable	Merkmalausprägungen	Referenzkategorie
Institution	0 = Institution P 1 = Institution Q	

Tab. 28: Lehr- und Lernumgebungsvariablen

5.3.2.2 Unterrichtsbezogene Variablen

Die Variable *Kurslevel* zeigt an, in welchem Kurslevel die Teilnehmenden eingestuft wurden, d. h., ob ihr Ergebnis beim Einstufungstest im oberen, mittleren oder unteren Drittel lag. Aus der Variable *Silbenmethode* wird ersichtlich, ob die Teilnehmenden in einem Kurs mit zusätzlichem Fokus auf der Silbenmethode unterrichtet wurden. Auch der Umstand, dass einige Teilnehmende bereits vor Besuch der Studie einen *Deutschkurs besucht* haben, wird mittels einer eigenen Variable abgebildet (s. Tab. 29).

Variable	Merkmalausprägungen	Referenzkategorie
Kurslevel	hoch:	0 = nein
		1 = ja
	mittel:	0 = nein
		1 = ja
	tief:	0 = nein
		1 = ja
Silbenmethode		0 = nein
		1 = ja
Besucher Deutschkurs		0 = nein
		1 = ja

Tab. 29: Unterrichtsbezogene Variablen

5.3.2.3 Personenbezogene Variablen

Die personenbezogenen Variablen bilden die grösste Variablengruppe und wurden mehrheitlich durch die von den Teilnehmenden ausgefüllten Fragebögen ermittelt (s. Tab. 30). Ausser den Kenntnissen in der Erstschriftsprache (*Wortschatz* und *formale Richtigkeit*) und der *Kursabsenz* handelt es sich bei allen Variablen um selbst gemachte Angaben durch die Teilnehmenden. Diese wurden nicht anderweitig überprüft.

Die Variablen *Geschlecht* ist selbsterklärend und konnte direkt aus dem Fragebogen entnommen werden.

Ein bekanntes Problem bei der *Altersangabe* von Flüchtlingen ist, dass ihnen ihr Geburtsdatum vielfach nicht oder nur ungefähr bekannt ist. Deshalb wird bei der Einreise in die Schweiz von der zuständigen Behörde oft der 1. Januar des vermuteten Geburtsjahrs als Geburtstag festgelegt. Im Fragebogen der vorliegenden Studie wurde deshalb nur nach dem Geburtsjahr und nicht nach dem genauen Datum gefragt. Um das Alter zu berechnen, wurde diese Variable so konstruiert, dass alle Teilnehmenden im Jahr 2018 ein Jahr älter wurden, d. h. der Stichtag war der 31.12.2018.

Wie viele Schuljahre die Probanden in ihrem Heimatland absolviert haben, wird mit der Variable *Schuljahre* angegeben. Die Variablen *Berufslehre*, *Universität* und *Arbeit Heimatland* wurden als Dummy-Variablen kodiert und jeweils mit dem Wert 1 versehen, wenn eine Person

über eine Berufslehre bzw. einen tertiären Abschluss verfügt oder in ihrem Heimatland vor der Flucht einer Arbeit nachgegangen ist. Ob die Teilnehmenden *Kinder* im betreuungspflichtigen Alter haben, wurde ebenfalls ermittelt. Zu Beginn des Projekts wurde im Fragebogen zudem nach der subjektiven Einschätzung des eigenen *Gesundheitszustands* gefragt. Dies wurde dann als mehrstufige Variable kodiert. Im Fragebogen wurde ausserdem die Frage gestellt, an wie vielen Tagen pro Woche die Teilnehmenden in ihrem Alltag Deutsch benutzen. Diese Angaben wurden dann in Clustern für ein- bis dreimal pro Woche und vier- bis siebenmal pro Woche zusammengefügt.

Auch die Erstsprachenkenntnisse bzw. die Erstschriftkenntnisse wurden mittels Fragebogen ermittelt. Bei der *Erstsprache (L1)* handelt es sich um die Sprache, die von den Probanden selbst als ihre Erstsprache angegeben wurde. Wie in Abschnitt 3.3.2 ausgeführt, korrelieren Erstsprache und Herkunftsland deutlich. Da dies zu fast identischen deskriptiven Beschreibungen, im Falle der Regression jedoch gar zu Problemen der Multikollinearität führen würde, wird in der Analyse nur auf die Erstsprache eingegangen.

Bei der Wortschatzbeherrschung und der formalen Richtigkeit in der Erstschrift (die von der Erstsprache abweichen kann, s. Abschn. 1.1.2) handelt es sich, im Gegensatz zu den anderen in diesem Abschnitt besprochenen Variablen, um eine Fremdbeurteilung, die durch Übersetzer erfolgte (s. Abschn. 4.3.2).

Als einzige personenbezogene Variable wurde die *Kursabsenz* nicht über den Fragebogen erhoben. Die Werte für diese kontinuierliche metrische Variable, die die Anzahl verpasster Kurslektionen der Teilnehmenden angibt, wurde den Absenzenblätter der zuständigen Lehrpersonen entnommen.

Variable	Merkmalausprägungen		Referenzkategorie
Geschlecht		0 = Mann 1 = Frau	
Alter	19 Jahre: 20–29 Jahre: 30–39 Jahre: 40–49 Jahre: über 50 Jahre:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	20–29 Jahre
Erstsprache (L1)	Tigrinya: Saho Arabisch: Kurdisch: Persisch: Tamil: Tibetisch:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	Tigrinya
Wortschatzbeherrschung Erstschrift	A2: B1/B2: C1/C2:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	A2
Formale Richtigkeit Erstschrift	A2: B1/B2: C1/C2:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	A2
Fremdsprache(n)		0 = nein 1 = ja	
Schuljahre	0 Jahre: 1–3 Jahre: 4–6 Jahre: 7–9 Jahre: 10–12 Jahre:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	7–9
Berufslehre		0 = nein 1 = ja	
Universität		0 = nein 1 = ja	
Arbeit Heimatland		0 = nein 1 = ja	
Kinder		0 = nein 1 = ja	
Gesundheitszustand	sehr schlecht: nicht gut: mittelmässig: ziemlich gut: sehr gut:	1 2 3 4 5	
Deutschgebrauch	nie: regemässig: häufig:	0 = nein 1 = ja 0 = nein 1 = ja 0 = nein 1 = ja	nie
Kursabsenz	Anzahl gefehlte Kurslektionen		

Tab. 30: Personenbezogene Variablen

5.3.3 Dummy-Variablen und Referenzkategorien

Ordinale und nominale Variablen wurden allesamt als Dummy-Variablen kodiert, damit sie in die lineare Regression aufgenommen werden konnten. Das Kurslevel der Kursteilnehmenden wurde zum Beispiel nicht anhand einer ordinalen Skala mit 1 = tief, 2 = mittel und 3 = hoch kodiert, sondern für jedes Kurslevel wurde eine separate Dummy-Variable gebildet, die die Merkmalsausprägung 0 für unzutreffend und 1 für zutreffend beinhaltet. Eine Teilnehmerin im tiefen Kurslevel weist somit den Wert 0 für die Dummy-Variablen mittel und hoch auf und den Wert 1 für die Variable *Kurslevel tief*. Im soeben beschriebenen Fall lässt sich beispielsweise das Kurslevel tief eindeutig durch die Ausprägungen *Kurslevel mittel* = 0 und *Kurslevel hoch* = 0 darstellen. Eine der drei Variablen ist daher nicht nur redundant, sondern führt im Falle einer linearen Regression zu Problemen der Multikollinearität. Aus diesem Grund muss bei einer linearen Regression immer eine Dummy-Variable ausgelassen werden. Diese Variable gilt als Referenzkategorie. Das bedeutet, dass die Korrelationskoeffizienten der korrespondierenden Variablen in der linearen Regression immer in Bezug auf diese Variable zu lesen sind. Bei einer linearen Regression auf das telc-Resultat als abhängige Variable bedeutet ein Regressionskoeffizient der Variable *Kurslevel mittel* von $B = 9.02$ entsprechend, dass Personen aus dem mittleren Kurslevel im Testresultat unter ansonsten gleichbleibenden Merkmalsausprägungen 9.02 Punkte mehr als das tiefe Kurslevel erreichten.

In der vorliegenden Arbeit werden zudem metrische Variablen, d. h. das Alter der Kursteilnehmenden und die Schulerfahrung in Jahren, als Dummy-Variablen operationalisiert. Dies soll auch nicht lineare Zusammenhänge aufdecken und dient, aus Sicht der Forscherin, der besseren Lesbarkeit der Resultate.

5.3.4 Stör- und Kontrollvariablen

Stör- resp. Drittvariablen sind in der Feldforschung kaum auszuschliessen, da es sich nicht um Laborsituationen handelt. Aus diesem Grund ist es entscheidend, sich einer möglichst umfassenden Zahl potenzieller Störfaktoren bewusst zu sein, um diese entweder zu vermeiden oder kontrollieren zu können. Dies soll sicherstellen, dass die gefundenen Ergebnisse einer Studie auch tatsächlich durch die untersuchten Variablen und nicht durch unbekannte Drittvariablen hervorgerufen werden (vgl. Albert/Marx 2016, S. 86).

Im Folgenden werden die Drittvariablen aufgelistet, die in dieser Arbeit berücksichtigt wurden:

1. Lehrpersonenbezogene Drittvariablen
 - a. Einstellungen der Lehrpersonen

- b. Bildung und beruflicher Hintergrund der Lehrpersonen
- c. Individueller Unterrichtsstil

2. Umweltbezogene Drittvariablen

3. Teilnehmerbezogene Drittvariablen

Um den Einfluss dieser Faktoren möglichst gering zu halten respektive auszuschliessen, wurden Kontrollvariablen konstruiert. Damit sind diejenigen Variablen gemeint, die konstant gehalten werden können, um einen unkontrollierten Einfluss auf die abhängigen Variablen zu vermeiden. So wurden **Punkt 1a und Punkt 1b** anhand eines Fragebogens erhoben, der unter anderem den Bildungs- und Erfahrungshintergrund und die Meinung der Lehrpersonen zum Kurskonzept erfragte (s. Abschn. 4.3.1). Zusätzlich fanden insgesamt zwei Austauschsitzen mit den Kursleitenden statt, an denen Chancen und Schwierigkeiten der Kurskonzepte besprochen wurden und die Einstellung der beteiligten Kurspersonen ersichtlich wurde (s. Abschn. 8.2). Der Unterrichtsstil der jeweiligen Lehrperson, **Punkt 1c**, wurde, soweit möglich, anhand der Semesterplanung berücksichtigt. Die Lehrpersonen mussten diese der Forscherin bei Kursende aushändigen. Individuelle didaktische Schwerpunkte lassen sich aus diesen Plänen jedoch kaum entnehmen und auch die Qualität des Unterrichts kann daraus nicht abgeleitet werden. Die einzige Möglichkeit, beides zu erfassen, wären systematische Hospitationen gewesen. Darauf wurde aber bewusst verzichtet, da aus eigener Erfahrung die Anwesenheit einer externen Person den Deutschunterricht stören kann. So ist zu beobachten, dass sich Teilnehmende durch Externe abgelenkt fühlen und dem Unterricht weniger konzentriert folgen. Zumindest aber für die Qualität des Unterrichts lassen sich die Berufserfahrung und der Bildungshintergrund der Kursleitenden als Proxy heranziehen.

Beide Institutionen waren zudem während des Pilotprojekts durch nicht kontrollierbare Ausseneinflüsse in Form deutlich wahrnehmbarer Lärmemissionen betroffen. Ob dies den Unterricht negativ beeinflusst hat, kann nicht beurteilt werden. Da diese Umwelteinflüsse in beiden Fällen vergleichbar waren, sollte sich **Punkt 2** deswegen nicht zulasten einer der beiden Institutionen ausgewirkt haben. Darüber hinaus wurden keine weiteren Umweltvariablen kontrolliert.

Es ist davon auszugehen, dass **Punkt 3**, die teilnehmerbezogenen Störvariablen, einen ausgeprägten Einfluss auf die Resultate hat. Deshalb wurden einige dieser Variablen in erklärende Variablen umgewandelt, um deren Einfluss zu kontrollieren, wie beispielsweise die Kursabsenz oder der Gesundheitszustand der Teilnehmenden. Es wären aber etliche weitere

Einflussfaktoren denkbar, die hier nicht überprüft wurden, wie die Selbsteinschätzung des eigenen Leistungsniveaus, die Motivation oder die Erwartung und Einstellung des Lernenden (vgl. Hattie 2013).

5.4 Datenaufbereitung

Die Datenaufbereitung und anschliessende Datenauswertung erfolgte mit Hilfe des Statistikprogramms SPSS 25. In einem ersten Schritt wurden alle Daten gemäss der oben besprochenen Kodierung manuell in die Software eingegeben. In einem zweiten Schritt wurde der Datensatz bereinigt. Teilnehmende, die nicht alle vier Tests absolviert haben, d. h. Personen, die an einer oder an beiden Follow-up-Messungen abwesend waren oder die telc-Prüfung nicht absolviert haben, wurden aus dem Datensatz entfernt. Dies traf auf fünf Personen zu. Kursteilnehmende, die für eine oder mehrere andere Variablen keine oder ungültige Werte aufgewiesen hätten, wären ebenfalls nicht in die Analyse aufgenommen worden. Da die Studienteilnehmenden aber beim Ausfüllen des Fragebogens betreut und direkt vor Ort auf fehlende Angaben aufmerksam gemacht wurden, waren keine solchen Lücken zu verzeichnen. Von den ursprünglich 62 Kursteilnehmenden (s. Abschn. 3.3.2) konnten nach der Bereinigung des Datensatzes 57 in die Datenauswertung aufgenommen werden.

Teil III: Ergebnisse

6. Analyse

Im Folgenden wird zuerst auf die Testgüte eingegangen (s. Abschn. 6.1) und anschliessend die Analysen der telc-Ergebnisse (s. Abschn. 6.2) und des Gesamtlernerfolgs (s. Abschn. 6.3) betrachtet. Zur Überprüfung der beiden Forschungsfragen werden in diesen beiden Abschnitten verschiedene statistische Analysen durchgeführt. In Abschnitt 6.2.3 werden die Anzahl erfolgreich absolvierter telc-A1-Prüfungen und die dabei erreichten Punkte als abhängige Variable gewählt. In Abschnitt 6.3.4 werden dieselben Tests mit dem Gesamtlernerfolg als abhängige Variable durchgeführt. Unabhängige Variablen sind jeweils die in Abschnitt 5.3.2 besprochenen Lehr- und Lernumgebungsvariablen, unterrichtsbezogenen Variablen und personenbezogenen Variablen.

6.1 Testgüte

Die Betrachtung und Berechnung der Testgüte geschieht anhand einer klassischen Itemanalyse, wie sie z. B. im Handbuch zur Entwicklung und Durchführung von Sprachtests der Association of Language Testers in Europe (ALTE) zu finden ist (vgl. ALTE 2012, S. 95–96). Damit wird die Güte der forschergenerierten Tests in Bezug zur ALTE-zertifizierten telc-A1-Prüfung gesetzt, die dieselben Analysen durchläuft.

Die vorliegende Itemanalyse wurde nach der Haupttestung mit allen Teilnehmenden durchgeführt, sodass jedes Item 57-mal bearbeitet wurde. Eine Pilotierung eines Itempools mit anschliessendem Ausschluss von ungeeigneten Items wurde in dieser Forschung nicht angestrebt, einerseits weil die Erprobung der Tests ohnehin mit einer kleinen Stichprobe durchgeführt werden musste (s. Abschn. 3.3), und andererseits, weil die Forschungsarbeit nicht zum Ziel hat, ein geeignetes Testverfahren für die Zielgruppe zu entwickeln.

Objektivität	Reliabilität	Validität
<ul style="list-style-type: none">•Durchführungsobjektivität•Beurteilungsobjektivität•Interpretationsobjektivität	<ul style="list-style-type: none">•Interne Konsistenz<ul style="list-style-type: none">•Schwierigkeitsindex•Trennschärfe•Retest-Reliabilität•Bewerter-Reliabilität	<ul style="list-style-type: none">•Konstruktvalidität•Inhaltsvalidität•Kriteriumsvalidität

Abb. 10: Objektivität, Reliabilität und Validität der forschergenerierten Tests

Im Folgenden wird auf die drei Gütekriterien *Objektivität*, *Reliabilität* und *Validität* eingegangen resp. auf Überlegungen und Massnahmen, um diesen Kriterien Rechnung zu tragen (s. Abb. 10).

6.1.1 Objektivität

Die Objektivität setzt die Unabhängigkeit der Ergebnisse von der Untersuchungsperson voraus. Testdurchführung, -auswertung und -interpretation müssen demzufolge standardisiert und unabhängig vom Testleiter erfolgen. Zu unterscheiden sind folglich die Durchführungsobjektivität, die Beurteilungsobjektivität und die Interpretationsobjektivität (vgl. Bühner 2011, S. 58).

Damit die Durchführungsobjektivität zu allen Messzeitpunkten – Einstufung, Follow-up 1 und Follow-up 2 – gewährleistet ist, wurden alle Aufgaben in den Tests standardisiert durchgeführt. Die Aufgaben in den Tests und die Anordnung der Testblätter waren jeweils identisch und unterschieden sich lediglich hinsichtlich der verwendeten Items. Die Durchführung erfolgte stets mit Hilfe von Audiodateien, die gleich lange Pausen und vorgängig Beispiele für die Teilnehmenden beinhalteten. Die Beispiele wurden entweder gemeinsam mit dem Prüfer oder im Falle des Diktats und des Tests zum Hörverstehen in der Gruppe mit der Forscherin betrachtet, bevor mit dem eigentlichen Test begonnen wurde. Die Vereinheitlichung des Testverfahrens und der Einsatz von Beispielen garantierten die gleiche Ausgangslage für alle Teilnehmenden und stellten sicher, dass alle Geprüften den Test verstanden hatten. Alle Tests wurden so einfach wie möglich gestaltet, sodass auf den schriftlichen Testblättern nur wenige zusätzliche Informationen enthalten waren. Die schriftliche Instruktion wurde zusätzlich mündlich durch die Testleiterin vorgelesen, damit alle Kursteilnehmenden die gleichen Arbeitsanweisungen erhielten und das Verständnis der Instruktion nicht an mangelndem Leseverständnis scheitern konnte. Die Kursteilnehmenden durften bei Verständnisschwierigkeiten zudem Fragen stellen.

Die Beurteilung der Tests erfolgte stets durch eine erfahrene Lehrperson und durch die Forscherin selbst. Dies geschah anhand der jeweiligen Beurteilungsraster und Kriterien (s. Abschn. 4.2.1.3 für EI, 4.2.2.3 für Kunstwörter, 4.2.3.3 für ORF, 4.2.4.3 für HV und 4.2.5.3 für Diktat). Wie in Abschnitt 4.2 erwähnt, erhielten die Bewerter vorgängig eine Schulung, in der das Evaluationsraster erklärt und allfällige Fragen beantwortet wurden. Anschliessend beurteilten alle Bewertenden zu jedem der vier Tests (HV ausgeschlossen) einige Testitems, die verglichen und diskutiert wurden und die anschliessend als Standards dienten.

Die Interpretationsobjektivität beschreibt «das Ausmass der Übereinstimmung in der Interpretation aufgrund gleicher Testergebnisse» (Bühner/Bernd 2009, S. 170). Sie ist dadurch gegeben, dass die beiden (Haupt-)Bewerterinnen bei den Beurteilungen der von den Kursteilnehmenden in den Tests gegebenen Antworten zu gleichen Ergebnissen gekommen sind und andernfalls ein Drittprüfer hinzugezogen wurde.

6.1.2 Reliabilität

Unter Reliabilität wird die Zuverlässigkeit der Leistungsmessung verstanden. Sie gibt die Messgenauigkeit des untersuchten Messwerts an (vgl. Bühner 2011, S. 60), wobei wie bei der Objektivität einzelne Aspekte zu unterscheiden sind. Im Weiteren wird auf folgende Kriterien eingegangen: interne Konsistenz (inklusive Schwierigkeitsindex und Trennschärfe), Retest-Reliabilität und Bewerterreliabilität.

6.1.2.1 Interne Konsistenz

Um die Zuverlässigkeit der Items im Hinblick auf die einzelnen Tests als Ganzes beurteilen zu können, wird die interne Konsistenz analysiert. Die am häufigsten verwendete Bestimmung dafür ist Cronbachs α , welche das Ausmass angibt, inwieweit die Items dasselbe Konstrukt messen und miteinander übereinstimmen. Ein Wert für Cronbachs α über .90 gilt als hoch (vgl. Bortz/Döring 2006, S. 199). Für tiefere Werte finden sich unterschiedliche Interpretationen, fest steht aber, dass ein Wert unter .70 als tief zu betrachten ist. Je höher α ist, desto grösser ist die interne Konsistenz der Items eines Tests (vgl. ebd.; Bühner/Bernd 2009, S. 175–177).

Um zu erfahren, «wie gut ein einzelnes Item das Gesamtergebnis eines Tests repräsentiert» (Bortz/Döring 2006, S. 219), wird die Trennschärfe ($r_{i,t}$) analysiert. Teilnehmer, die ein hohes Gesamtergebnis erreichen, sollten demnach auch bei trennscharfen Items einen hohen Einzelwert erhalten; für schwache Teilnehmer gilt das Umgekehrte. Bei der Trennschärfe sind keine festen Untergrenzen festgelegt, generell sind aber Trennschärfen mit einem Wert über .20 wünschenswert (vgl. Asendorpf/Neyer 2012, S. 54). Da wie eingangs erwähnt mit den forschergenerierten Tests nicht primär ein Testverfahren für die Zielgruppe konstruiert werden sollte, sondern vielmehr die Unterscheidung der Kursgruppen im Vordergrund steht, wurde gemäss Bühner und Bernd (2009, S. 79) die liberale Untergrenze von .00 gesetzt. Nur Items mit Werten im Minusbereich wurden somit aus den weiteren statistischen Analysen ausgeschlossen.

Um zudem Auskunft darüber zu erhalten, wie oft die Items richtig oder falsch gelöst wurden, wird der Schwierigkeitsindex ($p_{i,t}$) herangezogen. Der Schwierigkeitsgrad eines Items wird

hierbei durch die prozentuale Anzahl der Probanden definiert, die das entsprechende Item korrekt zu lösen vermochten, weshalb auch von relativer Lösungshäufigkeit gesprochen wird. Der Schwierigkeitsgrad (p) sollte laut Bühner (2011, S. 81) nicht unter .20 (zu schwer) und nicht über .80 (zu leicht) liegen. Es muss hier angemerkt werden, dass zu schwierige Items (<.20) und zu leichte Items (>.80) dennoch für die statistische Analyse beibehalten werden, wenn die Trennschärfe einen angemessenen Wert aufweist und sich der Schwierigkeitswert nicht im Ausschlussbereich von unter .05 bzw. über .95 befindet, wie dies von Bernd und Bühner vorgeschlagen wird (2009, S. 65, 78–79).

Elicited Imitation

Der Elicited Imitation zeigt zu allen drei Messzeitpunkten eine relativ hohe interne Konsistenz (Einstufung $\alpha = .94$, Follow-up 1 $\alpha = .91$ und Follow-up 2 $\alpha = .87$).

Einstufung				Follow-up 1				Follow-up 2			
Item	α	p	r	Item	α	p	r	Item	α	p	r
E1	.94	.44	.41	F1E1	.91	.83	.50	F2E1	.87	.53	.52
E2	.94	.34	.44	F1E2	.92	.71	.40	F2E2	.87	.67	.60
E3	.94	.52	.66	F1E3	.91	.28	.51	F2E3	.88	.29	.04
E4	.94	.39	.58	F1E4	.92	.85	.21	F2E4	.88	.76	.34
E5	.94	.48	.57	F1E5	.91	.71	.67	F2E5	.88	.62	.36
E6	.94	.52	.63	F1E6	.91	.66	.66	F2E6	.87	.40	.43
E7	.94	.73	.69	F1E7	.91	.28	.74	F2E7	.87	.44	.58
E8	.94	.36	.50	F1E8	.91	.47	.58	F2E8	.87	.54	.53
E9	.94	.43	.70	F1E9	.92	.32	.42	F2E9	.87	.62	.63
E10	.94	.39	.72	F1E10	.91	.31	.49	F2E10	.87	.51	.51
E11	.94	.43	.57	F1E11	.91	.52	.48	F2E11	.87	.72	.41
E12	.94	.51	.71	F1E12	.91	.47	.69	F2E12	.87	.84	.44
E13	.94	.41	.62	F1E13	.91	.70	.63	F2E13	.87	.51	.62
E14	.94	.64	.76	F1E14	.91	.33	.60	F2E14	.88	.41	.16
E15	.94	.39	.64	F1E15	.92	.49	.23	F2E15	.87	.23	.34
E16	.94	.57	.66	F1E16	.91	.47	.52	F2E16	.87	.44	.53
E17	.94	.38	.60	F1E17	.91	.29	.64	F2E17	.87	.45	.42
E18	.94	.29	.63	F1E18	.91	.33	.60	F2E18	.88	.47	.23
E19	.94	.37	.43	F1E19	.91	.53	.58	F2E19	.87	.37	.45
E20	.94	.29	.65	F1E20	.92	.43	.48	F2E20	.87	.28	.54
E21	.94	.41	.71	F1E21	.92	.36	.24	F2E21	.88	.45	.37
E22	.94	.44	.50	F1E22	.92	.40	.29	F2E22	.88	.39	.35
E23	.94	.33	.59	F1E23	.91	.53	.73	F2E23	.87	.36	.58
E24	.94	.33	.50	F1E24	.91	.27	.49	F2E24	.88	.29	.39
E25	.94	.29	.60	F1E25	.91	.30	.62	F2E25	.87	.61	.50
E26	.94	.28	.66	F1E26	.91	.42	.60	F2E26	.88	.27	.13
E27	.94	.34	.53	F1E27	.92	.24	.34	F2E27	.88	.27	.28
E28	.94	.35	.55	F1E28	.92	.43	.25	F2E28	.87	.35	.49
E29	.94	.38	.50	F1E29	.92	.27	.34	F2E29	.87	.40	.54
E30	.94	.21	.57	F1E30	.92	.32	.37	F2E30	.88	.21	.27

N = 57

Abb. 11: Reliabilitätsanalyse Elicited Imitation
 α = Cronbachs α , p = Schwierigkeitsindex und r = Trennschärfe

Bei genauerer Betrachtung zeigt sich, dass die Trennschärfen in der zweiten Follow-up-Messung mit $r = .04$ bis $.63$ geringer ausfallen als bei der ersten Follow-up-Messung mit $r = .21$

bis .74 und bei der Einstufung mit $r = .41$ bis .76 (s. Abb. 11). Da sich aber keine Werte im Negativbereich befinden, wurden keine Items ausgeschlossen. Die Werte für den Schwierigkeitsindex befinden sich ebenfalls innerhalb der angesetzten Bandbreite.

Kunstwörter

Die Kunstwörter weisen ebenfalls zu allen drei Messzeitpunkten eine im festgelegten Bereich liegende interne Konsistenz auf (Einstufung $\alpha = .79$, Follow-up 1 $\alpha = .80$ und Follow-up 2 $\alpha = .82$).

Einstufung				Follow-up 1				Follow-up 2			
Item	α	p	r	Item	α	p	r	Item	α	p	r
P1	.79	.86	.36	F1P1	.81	.93	.07	F2P2	.82	.86	.29
P2	.79	.93	.23	F1P2	.81	.98	-.08	F2P2	.83	.56	.00
P3	.80	.67	.09	F1P3	.81	.63	.13	F2P3	.83	.97	.13
P4	.81	.91	-.15	F1P4	.80	.97	.27	F2P4	.83	.86	.16
P5	.79	.68	.34	F1P5	.80	.86	.38	F2P5	.82	.86	.25
P6	.79	.84	.30	F1P6	.81	.93	.12	F2P6	.83	.67	.13
P7	.80	.74	.16	F1P7	.80	.70	.38	F2P7	.82	.33	.26
P8	.79	.37	.28	F1P8	.81	.37	.17	F2P8	.82	.54	.37
P9	.77	.51	.70	F1P9	.79	.58	.49	F2P9	.82	.61	.43
P10	.80	.51	.13	F1P10	.80	.58	.31	F2P10	.83	.53	.14
P11	.78	.56	.44	F1P11	.80	.63	.36	F2P11	.82	.90	.32
P12	.79	.88	.24	F1P12	.81	.90	.18	F2P12	.82	.42	.26
P13	.79	.53	.32	F1P13	.80	.44	.43	F2P13	.82	.37	.36
P14	.78	.32	.47	F1P14	.80	.40	.40	F2P14	.82	.35	.34
P15	.78	.28	.51	F1P15	.79	.26	.50	F2P15	.82	.32	.40
P16	.79	.26	.31	F1P16	.80	.28	.33	F2P16	.81	.37	.53
P17	.77	.32	.61	F1P17	.79	.33	.48	F2P17	.81	.25	.60
P18	.80	.12	.14	F1P18	.80	.09	.30	F2P18	.82	.11	.36
P19	.79	.11	.36	F1P19	.81	.07	.08	F2P19	.81	.12	.57
P20	.79	.07	.37	F1P20	.80	.07	.44	F2P20	.81	.12	.52
P21	.79	.14	.41	F1P21	.80	.07	.50	F2P21	.82	.18	.42
P22	.79	.21	.41	F1P22	.79	.19	.52	F2P22	.81	.09	.52
P23	.79	.11	.43	F1P23	.80	.04	.40	F2P23	.82	.10	.39
P24	.79	.04	.37	F1P24	.79	.16	.51	F2P24	.82	.07	.36
P25	.79	.04	.21	F1P25	.80	.05	.40	F2P25	.82	.04	.52
P26	.79	.04	.37	F1P26	.81	.02	.14	F2P26	.82	.05	.31
P27	.80	.02	.00	F1P27	.80	.04	.30	F2P27	.82	.07	.39
P28	.79	.02	.26	F1P28	.80	.05	.42	F2P28	.82	.05	.50
P29	.79	.02	.26	F1P29	.80	.04	.30	F2P29	.82	.02	.48
P30	.79	.86	.26	F1P30	.80	.04	.28	F2P30	.82	.86	-.05

N = 57

Abb. 12: Reliabilitätsanalyse Kunstwörter
 α = Cronbachs α , p = Schwierigkeitsindex und r = Trennschärfe

Bei genauerer Betrachtung von Abb. 12 zeigt sich, dass insgesamt drei Items einen negativen Trennschärfewert aufweisen. Das heisst, dass bei diesen Items schwächere Teilnehmende besser abschnitten als stärkere. Diese Items wurden entsprechend aus den Analysen ausgeschlossen ($r_{4,E}$, $r_{2,F1}$ und $r_{30,F2}$). Zu allen drei Messzeitpunkten ist zu beobachten, dass gemäss Schwierigkeitsindizes ab Kunstwort 18 die Items fast durchgehend als zu schwierig

eingestuft werden müssen. Dies könnte darauf zurückzuführen sein, dass ab Kunstwort 18 vier- bis sechssilbige Kunstwörter nachgesprochen werden mussten, weshalb der Schwierigkeitsgrad zwangsläufig merklich zunimmt.

ORF

Die Oral Reading Fluency kann keiner klassischen Itemanalyse unterzogen werden, weil das Ergebnis der jeweiligen Teilnehmenden lediglich aus einer einzelnen metrischen Variable besteht.

HV

Die Werte der internen Konsistenz befinden sich für alle drei Hörverstehenstests im definierten Bereich (Einstufung $\alpha = .89$, Follow-up 1 $\alpha = .76$ und Follow-up 2 $\alpha = .75$).

Einstufung				Follow-up 1			
Item	α	p	r	Item	α	p	r
E Hörverstehen 1	.90	.74	.42	F1 Hörverstehen 1	.78	.56	.10
E Hörverstehen 1	.89	.70	.56	F1 Hörverstehen 1	.75	.26	.40
E Hörverstehen 1	.89	.63	.74	F1 Hörverstehen 1	.76	.46	.30
E Hörverstehen 1	.89	.70	.54	F1 Hörverstehen 1	.76	.61	.30
E Hörverstehen 1	.89	.54	.66	F1 Hörverstehen 1	.78	.46	.14
E Hörverstehen 2	.89	.60	.64	F1 Hörverstehen 1	.77	.26	.11
E Hörverstehen 2	.89	.77	.54	F1 Hörverstehen 2	.75	.90	.47
E Hörverstehen 2	.89	.53	.63	F1 Hörverstehen 2	.75	.86	.44
E Hörverstehen 2	.89	.46	.72	F1 Hörverstehen 2	.74	.74	.53
E Hörverstehen 2	.89	.65	.74	F1 Hörverstehen 2	.74	.74	.61
E Hörverstehen 3	.89	.70	.57	F1 Hörverstehen 2	.75	.84	.44
E Hörverstehen 3	.90	.58	.34	F1 Hörverstehen 3	.76	.84	.41
E Hörverstehen 3	.89	.65	.53	F1 Hörverstehen 3	.78	.65	.12
E Hörverstehen 3	.89	.61	.47	F1 Hörverstehen 3	.77	.68	.18
E Hörverstehen 3	.90	.54	.31	F1 Hörverstehen 3	.76	.88	.25
E Hörverstehen 3	.89	.65	.65	F1 Hörverstehen 3	.74	.63	.57
E Hörverstehen 3	.90	.53	.39	F1 Hörverstehen 3	.76	.88	.41
E Hörverstehen 3	.90	.44	.30	F1 Hörverstehen 3	.75	.65	.48
				F1 Hörverstehen 3	.75	.68	.44

Follow-up 2			
Item	α	p	r
F2 Hörverstehen 1	.77	.63	.11
F2 Hörverstehen 1	.77	.84	.01
F2 Hörverstehen 1	.75	.33	.31
F2 Hörverstehen 1	.75	.60	.38
F2 Hörverstehen 1	.77	.54	.18
F2 Hörverstehen 1	.74	.49	.47
F2 Hörverstehen 2	.75	.95	.48
F2 Hörverstehen 2	.74	.90	.50
F2 Hörverstehen 2	.72	.72	.67
F2 Hörverstehen 2	.73	.81	.53
F2 Hörverstehen 2	.73	.77	.61
F2 Hörverstehen 3	.77	.95	.02
F2 Hörverstehen 3	.77	.60	.48
F2 Hörverstehen 3	.76	.91	.14
F2 Hörverstehen 3	.76	.91	.20
F2 Hörverstehen 3	.76	.72	.25
F2 Hörverstehen 3	.75	.91	.27
F2 Hörverstehen 3	.76	.84	.22
F2 Hörverstehen 3	.74	.74	.46

N = 57

Abb. 13: Reliabilitätsanalyse Hörverstehen
 α = Cronbachs α , p = Schwierigkeitsindex und r = Trennschärfe

Bei Betrachtung der Abb. 13 zeigt sich, dass die Trennschärfen der zweiten Follow-up-Messung $r = .01$ bis $.67$ geringer ausfallen als bei der ersten Follow-up-Messung mit $r = .10$ bis $.61$ und bei der Einstufung $r = .30$ bis $.74$. Die Werte befinden sich aber auch hier noch im definierten Bereich, weshalb kein Item-Ausschluss vorgenommen wurde.

Die Auswertung des Tests zum Hörverstehen erfolgte dichotom: Das Item kann als korrekt (volle Punktzahl) oder als nicht korrekt (keine Punkte) bewertet werden, wobei keine Teilpunkte vergeben wurden. Der Schwierigkeitsindex kann demnach im Unterschied zum Elicited Imitation und den Kunstwörtern nur anhand des Mittelwerts betrachtet werden, denn dieser «drückt den relativen Anteil der Probanden aus, die das Item richtig beantwortet haben» (Bühner 2011, S. 243). Auch wenn einige Schwierigkeitsindizes in beiden Follow-up-Messungen als sehr hoch und dementsprechend als sehr leicht für die Teilnehmenden einzustufen sind, wurden sie dennoch nicht aus den Analysen ausgeschlossen, da sie sich nicht im Ausschlussbereich jenseits von $.05$ bzw. $.95$ befinden. Beide Tests als Ganze weisen zudem einen im definierten Bereich liegenden Schwierigkeitsindex auf ($p_{F1} = .63$ und $p_{F2} = .71$), was insgesamt für leichtere Hörverstehenstests mit akzeptabler Schwierigkeit spricht.

Diktat

Die interne Konsistenz im Diktat weist für alle drei Tests einen hohen Wert auf (Einstufung $\alpha = .92$, $F1 \alpha = .90$ und $F2 \alpha = .93$). Die Werte für die Trennschärfe und den Schwierigkeitsindex befinden sich innerhalb des definierten Wertbereichs (s. Abb. 14), weshalb alle Items in die Analysen aufgenommen wurden.

Einstufung				Follow-up 1				Follow-up 2			
Item	α	p	r	Item	α	p	r	Item	α	p	r
DB1	.92	.35	.66	F1DB 1	.90	.26	.36	F2DB1	.93	.74	.57
DB2	.92	.34	.45	F1DB 2	.90	.57	.34	F2DB2	.93	.51	.45
DB3	.92	.72	.31	F1DB 3	.91	.79	.17	F2DB3	.93	.64	.34
DB4	.92	.45	.46	F1DB 4	.90	.73	.54	F2DB4	.93	.62	.42
DB5	.92	.56	.54	F1DB 5	.90	.65	.46	F2DB5	.93	.49	.60
DB6	.92	.73	.56	F1DB 6	.90	.36	.57	F2DB6	.93	.71	.47
DB7	.92	.41	.41	F1DB 7	.90	.43	.54	F2DB7	.93	.48	.33
DB8	.92	.33	.40	F1DB 8	.90	.63	.61	F2DB8	.93	.75	.65
DB9	.92	.57	.56	F1DB 9	.90	.69	.48	F2DB9	.93	.41	.40
DB10	.92	.44	.64	F1DB 10	.90	.77	.67	F2DB10	.93	.54	.58
DB11	.92	.36	.30	F1DB 11	.90	.47	.51	F2DB11	.93	.30	.29
DB12	.92	.16	.46	F1DB 12	.90	.46	.44	F2DB12	.92	.52	.76
DB13	.92	.55	.55	F1DB 13	.90	.33	.28	F2DB13	.93	.75	.49
DB14	.92	.60	.63	F1DB 14	.90	.61	.43	F2DB14	.93	.64	.48
DB15	.92	.42	.50	F1DB 15	.90	.64	.53	F2DB15	.93	.68	.51
DB16	.92	.67	.40	F1DB 16	.90	.35	.28	F2DB16	.93	.85	.29
DB17	.92	.57	.43	F1DB 17	.90	.71	.45	F2DB17	.93	.56	.58
DB18	.92	.53	.58	F1DB 18	.90	.43	.54	F2DB18	.93	.65	.63
DB19	.92	.39	.53	F1DB 19	.90	.43	.35	F2DB19	.92	.60	.68
DB20	.92	.26	.66	F1DB20	.90	.54	.62	F2DB20	.93	.60	.64
DB21	.92	.18	.64	F1DB21	.90	.23	.54	F2DB21	.92	.52	.64
DB22	.92	.22	.64	F1DB22	.90	.20	.42	F2DB22	.93	.40	.60
DB23	.92	.36	.66	F1DB23	.90	.47	.36	F2DB23	.93	.27	.54
DB24	.92	.15	.57	F1DB24	.90	.19	.44	F2DB24	.93	.35	.64
DB25	.92	.16	.60	F1DB25	.90	.23	.63	F2DB25	.93	.34	.64
DB26	.92	.16	.49	F1DB26	.90	.16	.53	F2DB26	.92	.49	.71
DB27	.92	.19	.60	F1DB27	.90	.51	.72	F2DB27	.93	.34	.59
DB28	.92	.26	.73	F1DB28	.90	.28	.51	F2DB28	.92	.35	.69
DB29	.92	.15	.39	F1DB29	.90	.22	.58	F2DB29	.92	.54	.69
DB30	.92	.19	.50	F1DB30	.90	.17	.58	F2DB30	.93	.38	.68

N = 57

Abb. 14: Reliabilitätsanalyse Diktat
 α = Cronbachs α , p = Schwierigkeitsindex und r = Trennschärfe

6.1.2.2 Retest-Reliabilität

Mit der Retest-Reliabilität wird erhoben, ob ein Test über verschiedene Zeitpunkte hinweg stabile Messergebnisse erzeugt. Dazu wird die Korrelation zwischen zwei gleichen, aber zu unterschiedlichen Zeitpunkten durchgeführten Tests – oder wie in der vorliegenden Untersuchung zwischen identischen Testitems aus drei verschiedenen Tests – herangezogen. 'Stabile Messerzeugnisse' bedeuten hierbei nicht, dass die durch die Teilnehmenden erzielten Resultate bei den wiederholten Items gleich ausfallen; dies würde der Idee einer Lernprogression zwischen zwei Messzeitpunkten widersprechen. Stabile Werte implizieren vielmehr eine sehr hohe positive, nicht aber hundertprozentige Korrelation. Da die untersuchten Items deckungsgleich sind, kann nicht gänzlich ausgeschlossen werden, dass Erinnerungseffekte die Resultate künstlich in die Höhe treiben, weshalb auf die Wahl eines geeigneten zeitlichen Abstands zwischen beiden Messzeitpunkten geachtet werden muss (vgl.

Bühner/Bernd 2009, S. 174–175). Im vorliegenden Fall wurden nur einzelne Testitems wiederholt und zwischen den Messzeitpunkten lagen hundert Lektionen, weshalb von nicht sonderlich ausgeprägten Erinnerungseffekten auszugehen ist.

Bei der Elicited Imitation und dem Diktat wurden insgesamt 30 % der Testitems aus dem Einstufungstest in den beiden Follow-up-Messungen wiederholt. Beim ORF war insgesamt eine Lesezeile zu allen drei Messzeitpunkten identisch und beim Hörverstehen wurde jeweils die dritte Aufgabe, d. h. das detaillierte Hören, wiederholt. Die Kunstwörter wurden allesamt wiederholt, da die Wahrscheinlichkeit des Erinnerns gering ist, da es sich hier, um nicht reale und teilweise komplexe Wörter handelt.

Item	M	SD	Item	M	SD
E16	2.56	1.18	E7	2.93	1.35
<i>E14</i>	2.26	1.13	<i>E21</i>	2.30	1.07
E6	2.09	1.37	E3	2.07	1.24
E5	2.04	1.55	E11	1.74	.84
E1	1.93	1.18	E13	1.63	.86
<i>E12</i>	1.75	1.47	E28	1.40	.68
E22	1.74	1.06	E18	1.18	.93
E4	1.72	.75	<i>E26</i>	1.11	.79
<i>E9</i>	1.58	1.46	E30	.82	.66
E10	1.56	.87			
<i>E15</i>	1.54	.85			
E17	1.53	.68			
<i>E19</i>	1.51	.76			
E29	1.49	.71			
E2	1.42	1.07			
<i>E8</i>	1.37	1.41			
E27	1.35	.67			
E23	1.33	.81			
E24	1.33	.69			
<i>E20</i>	1.16	.77			
E25	1.16	.73			

Tab. 31: Mittelwert der Items, Elicited Imitation, Einstufung
Kursive Ziffer: Retest-Item

Zur Wahl der Retest-Items des Elicited Imitation wurden zuerst alle A1- und A2-Items des Einstufungstests getrennt und entlang der beim Einstufungstest erreichten durchschnittlichen Punktzahl sortiert (s. Tab. 31). Die 21 A1-Items wurden anschliessend aufgrund der durchschnittlichen Punktzahl in drei Gruppen zu je sieben Items eingeteilt. Aus der Gruppe mit den sieben schwierigeren und aus der Gruppe mit den sieben leichteren Items wurden je das zweiteinfachste und das zweitschwerste Item (E14 und E12, resp. E8 und E20) ausgewählt. Aus der mittleren Gruppe wurden drei Items – das zweiteinfachste, das zweitschwerste und

das mittlere Item (E9, E15, E19) – herausgezogen. Von den A2-Items wurden das zweitschwerste (E26) und das zweiteinfachste (E21) Item ausgewählt. Dies ergab ein Total von insgesamt neun Testitems, die als Retest-Items in die beiden Follow-up-Messungen aufgenommen wurden: E8 (Muss er der Strasse folgen?), E9 (Such das im Internet!), E12 (Ich habe dich getroffen.), E14 (Die Kinder helfen der Mutter.), E15 (Lernen Sie wenigstens fünf Wörter!), E19 (Meine Töchter sind unterschiedlich gross.), E20 (Haben Sie einen guten Eindruck von mir?), E21 (Beginnt deine neue Arbeit im August?) und E26 (Ich denke jeden Tag an unsere Vergangenheit.):

Da das Diktat einerseits anhand der Verständlichkeit (0–2 Punkte) und andererseits in Bezug auf die Laut-Buchstaben-Zuordnung (L-B-Z, 0–5 Punkte) bewertet wurde, wurde zuerst für jedes Item der kombinierte Mittelwert der beim Einstufungstest erreichten Punktzahl berechnet, indem die Punktezahl auf 10 normiert wurde: $[(\text{Punktzahl Verständlichkeit} * 5) \text{ plus } (\text{Punktzahl L-B-Z} * 2)]$ geteilt durch 2 = Mittelwert Diktat. Anschliessend wurde der Durchschnittswert aller A2-Sätze ermittelt (1.95) und derjenige Satz ausgewählt, der dem Durchschnittswert am nächsten kam: Satz D24 ($M = 1.87$, Ich entschuldige mich.). Für den Durchschnittswert aller A1-Sätze ($M = 2.09$) wurde das gleiche Prozedere durchgeführt und Satz D25 ($M = 1.63$, Die Arbeit macht Spass.) ausgewählt.

Item	M	SD	Item	M	SD
D6	7.27	2.95	D5	5.73	3.38
D3	7.24	3.29	D9	5.55	4.02
D14	6.04	3.79	D15	4.20	4.03
D17	5.74	3.99	D22	2.61	2.52
D18	5.54	3.81	D28 (Satz)	2.60	3.08
D13	5.28	3.84	D20	2.19	2.54
D4	4.49	3.67	D24 (Satz)	1.87	1.40
D7	4.39	3.73	D27 (Satz)	1.85	1.60
D10	4.14	3.86	D30 (Satz)	1.50	1.28
D19	3.87	3.11			
D11	3.68	3.32			
D16	3.62	2.88			
D23 (Satz)	3.61	3.09			
D1	3.47	3.91			
D2	3.36	3.27			
D8	3.28	2.50			
D21	1.82	2.37			
D25 (Satz)	1.63	1.65			
D26 (Satz)	1.61	2.49			
D12	1.58	2.66			
D29 (Satz)	1.53	1.26			

Tab. 32: Mittelwert der Items, Diktat, Einstufung
Kursiv: Retest-Item

Bei den A2-Wörtern wurden das zweiteinfachste und das zweitschwerste Wort als Retest-Items ausgewählt (s. Tab. 32), D9 (Winter) und D22 (Erwachsene). Unter den A1-Wörtern wurde entlang der aufsteigenden Schwierigkeit jedes dritte Item herausgezogen (s. Tab. 32), D8 (finden), D16 (stossen), D10 (einfach), D13 (teilen) und D14 (nochmal).

Zur Messung der Retest-Reliabilität wurde anschliessend für jeden Test und jeden Messzeitpunkt eine Variable mit der durchschnittlichen Punktzahl der jeweiligen Retest-Items in SPSS gebildet. Auf diese Weise konnte die Korrelation der Retest-Items über die drei Messzeitpunkte hinweg, d. h. von der Einstufung bis hin zur zweiten Follow-up-Messung, geprüft werden. Für Elicited Imitation (s. Tab. 33), das Diktat (s. Tab. 34) und die Kunstwörter (s. Tab. 35) ist durchgehend eine hohe Korrelation zwischen den drei Messzeitpunkten erkennbar. Die Werte liegen alle über dem von Bernd und Bühner (2009, S. 177) empfohlenen Schwellenwert von $> .60$ und sind somit ein Indiz für eine mindestens befriedigende Retest-Reliabilität.

		Einstufung	Follow-up 1	Follow-up 2
EI	Korrelation nach Pearson	1	.711**	.662**
	Signifikanz (2-seitig)		.000	.000
EI	Korrelation nach Pearson	.711**	1	.778**
	Signifikanz (2-seitig)	.000		.000
EI	Korrelation nach Pearson	.662**	.778**	1
	Signifikanz (2-seitig)	.000	.000	
N		57	57	57

Tab. 33: Elicited Imitation, Korrelation der Retest-Items

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

		Einstufung	Follow-up 1	Follow-up 2
Diktat	Korrelation nach Pearson	1	.771**	.758**
	Signifikanz (2-seitig)		.000	.000
Diktat	Korrelation nach Pearson	.771**	1	.824**
	Signifikanz (2-seitig)	.000		.000
Diktat	Korrelation nach Pearson	.758**	.824**	1
	Signifikanz (2-seitig)	.000	.000	
N		57	57	57

Tab. 34: Diktat, Korrelation der Retest-Items

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

		Einstufung	Follow-up 1	Follow-up 2
Kunstwörter	Korrelation nach Pearson	1	.820**	.818**
	Signifikanz (2-seitig)		.000	.000
Kunstwörter	Korrelation nach Pearson	.820**	1	.999**
	Signifikanz (2-seitig)	.000		.000
Kunstwörter	Korrelation nach Pearson	.818**	.999**	1
	Signifikanz (2-seitig)	.000	.000	
N		57	57	57

Tab. 35: Kunstwörter, Korrelation der Retest-Items

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

Beim Hörverstehen liegt die Korrelation zwischen dem Einstufungstest und der ersten Follow-up-Messung bei .627 (s. Tab. 36). Die zweite Follow-up-Messung korreliert mit .499 resp. .543 weniger ausgeprägt mit den anderen Messungen.

		Einstufung	Follow-up 1	Follow-up 2
HV	Korrelation nach Pearson	1	.627**	.499**
Einstufung	Signifikanz (2-seitig)		.000	.000
HV	Korrelation nach Pearson	.627**	1	.543**
Follow-up 1	Signifikanz (2-seitig)	.000		.000
HV	Korrelation nach Pearson	.499**	.543**	1
Follow-up 2	Signifikanz (2-seitig)	.000	.000	
N		57	57	57

Tab. 36: Hörverstehen, Korrelation der Retest-Items

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

6.1.2.3 Bewerter-Reliabilität

Zur Bewerter-Reliabilität gehören die Interrater-Reliabilität, die die Übereinstimmung zwischen den Bewertungen verschiedener Personen misst, und die Intrarater-Reliabilität, die die Übereinstimmung beim gleichen Beurteiler über verschiedene Zeitpunkte ermittelt (vgl. ALTE 2012, S. 99–100).

Der Elicited Imitation, die Kunstwörter und das Diktat wurden zum einen von einer Bewerterin und zum anderen von der Forscherin jeweils unabhängig bewertet. Die Bewertungen wurden mithilfe eines vorgegebenen Beurteilungsrasters (s. Abschn. 4.2.1.3, 4.2.2.3 und 4.2.5.3) vorgenommen und stimmten mehrheitlich überein. In den Fällen, in denen sich eine Diskrepanz ergab, wurde eine dritte Person herangezogen.

Für die Interrater-Reliabilität wird ein Intra-Class-Korrelationskoeffizient (ICC) von mindestens .60 gefordert, wenn keine folgenreiche Entscheidungen am Testergebnis hängen und es sich um eine Gruppenuntersuchung handelt, wobei alle Werte über .80 als exzellent gelten (vgl. Bühner/Bernd 2009, S. 177). Die Interrater-Reliabilität ist im Elicited Imitation ($ICC_E = .96$, $ICC_{F1} = .93$ und $ICC_{F2} = .90$) und im Diktat ($ICC_E = .95$, $ICC_{F1} = .94$ und $ICC_{F2} = .96$) als exzellent, bei den Kunstwörtern in der Einstufung ($ICC_E = .67$) als hoch und in beiden Follow-up-Messungen ($ICC_{F1} = .83$ und $ICC_{F2} = .86$) ebenfalls als exzellent zu beurteilen.

Bei der Intrarater-Reliabilität, die mittels Cohens Kappa in SPSS gemessen wird, gilt ebenfalls ein Wert über .80 als exzellent (vgl. Landis/Koch 1977, S. 165). Die Intrarater-Reliabilität ist beim Elicited Imitation bei beiden Ratern mit $IRR = .86$ als exzellent einzustufen, das Gleiche gilt für das Diktat ($IRR = .90$).

Die sehr hohen Intra-Class-Korrelationskoeffizienten und die ebenfalls sehr hohen Werte der Intrarater-Reliabilität lassen auf eine konsistente Beurteilung der Bewerter schliessen.

6.1.3 Validität

Die Validität ist ein Prüfkriterium für die Gültigkeit der Testverfahren und gibt die Eignung eines Messverfahrens für die Überprüfung der Fragestellung und der Zielsetzung an. Die Ergebnisse eines Tests gelten als valide, wenn diese einen «unmittelbaren und fehlerfreien Rückschluss» (Lienert/Raatz 1998, S. 10) auf die zu erfassenden Merkmale zulassen. Hier soll folglich ermittelt werden, ob die Tests das messen, was sie zu messen vorgeben. Dazu werden im folgenden drei grundlegende Kriterien näher beschrieben.

6.1.3.1 Konstruktvalidität

Ein Test als Konstrukt gilt als valide, wenn auf Basis der Theorie eine empirische Bestätigung der Hypothesen erfolgt (vgl. Bortz/Döring 2006, S. 201). Aus der Theorie oder der Empirie zu Zweitschriftlernenden kann indes nur wenig abgeleitet werden, da der Stand der Forschung dünn und nur ein kleiner Teil des Wissens empirisch gesichert ist.

Dies ist mitunter ein Grund, weshalb in Abschnitt 4.1 das Sprachverwendungsmodell für das forschergenerierte Testverfahren sowie die telc-Prüfung erläutert und vergleichend betrachtet werden. Diese klare Eingrenzung und Definition der Sprachverwendungsmodelle und der Testkonstrukte soll die Angemessenheit der Testinhalte sicherstellen (vgl. Bolton 2000, S. 12).

6.1.3.2 Inhaltsvalidität

Die Inhaltsvalidität lässt sich nicht numerisch berechnen, sondern muss subjektiv beurteilt werden. Dementsprechend handelt es sich hier weniger um ein Testgütekriterium, als vielmehr um eine Zielvorgabe (vgl. Bortz/Döring 2006, S. 200). Die unterschiedlichen Tests untersuchen jeweils ein bestimmtes Phänomen, was in Abschnitt 4.2.1 behandelt wurde. Anhand dieser Angaben kann davon ausgegangen werden, dass die Tests tatsächlich das messen, was sie sollen. Die Inhaltsvalidität kann hier unter Verweis auf die Testkonstruktion als gegeben betrachtet werden.

6.1.3.3 Kriteriumsvalidität

Die Kriteriumsvalidität misst «die Gültigkeit der Schlüsse aus einem Testergebnis auf ein praktisch relevantes Kriterium ausserhalb der Testsituation» (Bühner/Bernd 2009, S. 192). Die Kriteriumsvalidität wird in der Regel anhand der Korrelation zwischen den Test- und Kriteriumswerten einer Stichprobe geprüft (vgl. Bortz/Döring 2006, S. 200).

Obwohl die forschergenerierten Tests und die telc-Prüfung nicht das gleiche Sprachverwendungsmodell zur Grundlage haben, so messen sie doch beide einen

leistungsdifferenzierten Sprachstand (s. Abschn. 4.1). Dementsprechend werden in diesem Abschnitt die Tests auf korrelationsstatistischem Weg in Beziehung gesetzt (s. Tab. 37). Für die Interpretation der Korrelation werden die Grenzwerte nach Cohen (1988, S. 25–27) herangezogen, nach dem $r = .10$ für einen schwachen Effekt, $r = .30$ für einen mittleren und $r = .50$ für einen starken Effekt gelten.

		telc Hören	telc Lesen	telc Schreiben	telc Sprechen	telc Total
Einstufung	Korrelation nach Pearson	.343**	-.081	.235	.487**	.389**
Elicited Imitation	Signifikanz (2-seitig)	.009	.547	.078	.000	.003
Kunstwörter	Korrelation nach Pearson	.236*	.086	.198	.146	.264*
	Signifikanz (2-seitig)	.048	.523	.140	.277	.047
ORF	Korrelation nach Pearson	.186	.086	.294*	.429**	.384**
	Signifikanz (2-seitig)	.166	.525	.026	.001	.003
Hörverstehen	Korrelation nach Pearson	.473**	.182	.428**	.545**	.568**
	Signifikanz (2-seitig)	.000	.174	.001	.000	.000
Diktat	Korrelation nach Pearson	.198	-.180	.320*	.394**	.320*
	Signifikanz (2-seitig)	.141	.181	.015	.002	.015
Follow-up 1	Korrelation nach Pearson	.400**	.001	.235	.498**	.433**
Elicited Imitation	Signifikanz (2-seitig)	.002	.997	.079	.000	.001
Kunstwörter	Korrelation nach Pearson	.272*	.085	.200	.149	.268*
	Signifikanz (2-seitig)	.041	.529	.136	.269	.043
ORF	Korrelation nach Pearson	.186	.086	.294*	.429**	.384**
	Signifikanz (2-seitig)	.166	.525	.026	.001	.003
Hörverstehen	Korrelation nach Pearson	.473**	.240	.447**	.545**	.638**
	Signifikanz (2-seitig)	.000	.072	.000	.000	.000
Diktat	Korrelation nach Pearson	.242	-.118	.428**	.367**	.396**
	Signifikanz (2-seitig)	.070	.380	.001	.005	.002
Follow-up 2	Korrelation nach Pearson	.415**	-.040	.269*	.595**	.480**
Elicited Imitation	Signifikanz (2-seitig)	.001	.767	.043	.000	.000
Kunstwörter	Korrelation nach Pearson	.276*	.093	.200	.154	.274*
	Signifikanz (2-seitig)	.037	.494	.136	.252	.039
ORF	Korrelation nach Pearson	.257	.076	.285*	.404**	.391**
	Signifikanz (2-seitig)	.054	.577	.032	.002	.003
Hörverstehen	Korrelation nach Pearson	.333*	.222	.423**	.463**	.548**
	Signifikanz (2-seitig)	.011	.097	.001	.000	.000
Diktat	Korrelation nach Pearson	.370**	-.093	.380**	.429**	.444**
	Signifikanz (2-seitig)	.005	.491	.004	.001	.001

Tab. 37: Korrelation forschergenerierte Tests zu telc
* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

Der Elicited Imitation, der die Produktion in der Zweitsprache überprüft, korreliert wie erwartet mittel bis stark mit telc Sprechen ($r_E = .487^{**}$, $r_{F1} = .498^{**}$ und $r_{F2} = .595^{**}$). Auch das Diktat ($r_E = .320^*$, $r_{F1} = .428^{**}$ und $r_{F2} = .380^{**}$) korreliert mittelstark mit telc Schreiben, wenngleich im Diktat lediglich die Laut-Buchstaben-Zuordnung abgefragt wird und im telc-Test ein Brief geschrieben werden muss. Das Hörverstehen korreliert mittelstark mit telc Hören ($r_E = .473^{**}$, $r_{F1} = .473^{**}$ und $r_{F2} = .333^*$). Die Oral Reading Fluency zeigt überraschenderweise keine Korrelation mit telc Lesen ($r_E = .086$, $r_{F1} = .086$ und $r_{F2} = .076$). Dies

könnte darauf zurückzuführen sein, dass im forschergenerierten Test eher Bottom-up- und im telc Lesen eher Top-down-Aspekte gemessen werden.

Die Kunstwörter, welche eine Komponente des Arbeitsgedächtnisses messen, korrelieren zu allen drei Messzeitpunkten ($r_E = .264^*$, $r_{F1} = .268^*$ und $r_{F2} = .274^*$) eher schwach mit dem telc-Gesamtergebnis, was darauf schliessen lässt, dass die phonologische Schleife (s. Abschn. 4.2.2.1) für das Abschneiden bei der telc-Prüfung eine eher untergeordnete Rolle spielen könnte.

6.2 Deskriptive Analyse der telc-Prüfungsergebnisse

Voraussetzung für die Anwendung vertiefender statistischer Testverfahren ist eine Normalverteilung der intervallskalierten Variablen. Im Falle der telc-Prozentwerte bestätigt der Shapiro-Wilk-Test die Normalverteilung der Daten ($p = .178$, s. Abb. 15). Hinsichtlich der Anzahl der Teilnehmenden, die die telc-Prüfung bestanden haben, handelt es sich hingegen um eine kategoriale Variable (pass/fail), weshalb keine Annahmen über die Verteilungsform vorausgesetzt werden.

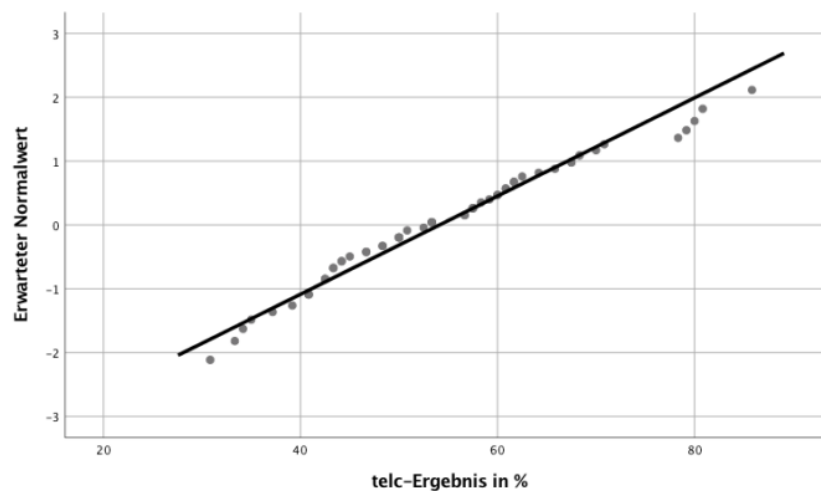


Abb. 15: telc-Prozentwerte (Punkte) und Normalverteilung (Linie)

Um die Aussagekraft statistischer Zusammenhänge genauer zu bestimmen, wird die Effektstärke nach Cohen (1988, S. 25–27) herangezogen (s. Tab. 38):

Effektgrösse	Schwacher Effekt	Mittlerer Effekt	Starker Effekt	Verwendung
d	.20	.50	.80	t-Test
η^2	.06	.10	.14	ANOVA
r	.10	.30	.50	Korrelation und Regression

Tab. 38: Effektstärke nach Cohen

Der unten stehenden Tab. 39 sind alle Variablen und statistischen Analysen zu entnehmen, auf die in Abschnitt 6.2.1 eingegangen wird.

Variable	Deskriptive Analyse		Lineare Regression (Erfolgsquote)
	Erfolgsquote telc	telc-Prüfungsergebnis	
Institution	Institution Q (45 %) > Institution P (19 %): $\chi^2(1) = 4.28$, $\Phi = .27$, ($p = .043$)	Institution Q ($M = 57\%$) > Institution P ($M = 50\%$): $t(55) = -2.07$, $p = .038$, $d = .56$	$B = 10.25$, $p = .000$
Kurslevel	nicht sign.	Kurslevel hoch ($M = 47\%$) > Kurslevel tief ($M = 61\%$): $F(2,54) = 6.8$, $p = .002$, partielles $\eta^2 = .20$	Kurslevel mittel: $B = 9.02$, $p = .003$ Kurslevel hoch: $B = 16.60$, $p = .000$
Geschlecht	nicht sign.	nicht sign.	$B = 7.98$, $p = .015$
Alter	nicht sign.	nicht sign.	Alter ≤ 19 : $B = -2.06$, $p = .017$ Alter 30 bis 39: $B = 14.20$, $p = .000$
Erstsprache (L1)	nicht sign.	Erstsprache: $F(5,50) = 2.96$, $p = .015$, partielles $\eta^2 = .26$	L1 Arabisch: $B = 14.66$, $p = .002$ L1 Persisch: $B = -17.91$, $p = .028$
Wortschatzbeherrschung Erstschrift	nicht sign.	nicht sign.	nicht sign.
Formale Richtigkeit Erstschrift	nicht sign.	nicht sign.	nicht sign.
Fremdsprache	nicht sign.	nicht sign.	nicht sign.
Anzahl Schuljahre	nicht sign.	nicht sign.	4 bis 6 Schuljahre: $B = -8.51$, $p = .007$ 10 bis 12 Schuljahre: $B = 8.43$, $p = .008$
Arbeit Heimatland	nicht sign.	nicht sign.	$B = -6.65$, $p = .009$
Berufslehre	nicht sign.	nicht sign.	$B = -8.01$, $p = .032$
Universität	nicht sign.	nicht sign.	nicht sign.
Deutschgebrauch	nicht sign.	nicht sign.	1–3x pro Woche: $B = 16.04$, $p = .003$ 4–7x pro Woche: $B = 16.80$, $p = .001$
Besucher Deutschkurs	Kein Deutschkurs (64 %) > besuchter Deutschkurs (26 %): $\chi^2(1) = 5.63$, $\Phi = -.31$, $p = .022$	Kein Deutschkurs ($M = 62\%$) > besuchter Deutschkurs ($M = 52\%$): $t(55) = 2.37$, $p = .022$, $d = .65$	$B = -21.70$, $p = .000$
Silbenmethode	nicht sign.	Kurslevel tief mit Silbenmethode ($M = 52\%$) > Kurslevel tief ohne Silbenmethode ($M = 42\%$): $t(45) = -2.253$, $p = .038$, $d = 1.0$	$B = 5.02$, $p = .039$

Tab. 39: Übersichtstabelle statistische Analysen telc
Erfolgsquote telc: Prozentsatz bestandener telc-Prüfungen
telc-Prüfungsergebnis: durchschnittlich erreichte Punkte in Prozent

Im Fliesstext werden alle Variablen einzeln betrachtet, auch wenn die statistischen Tests keine signifikanten Werte aufweisen. Da die Mehrheit der Variablen theorie- und literaturgeleitet

oder aufgrund persönlicher Praxiserfahrung gebildet wurde, ist auch das Fehlen signifikanter Werte, d. h. das Ausbleiben eines überprüfbaren Zusammenhangs, aufschlussreich. Zudem kann nicht ausgeschlossen werden, dass bei Verwendung einer grösseren Stichprobe einige Variablen doch signifikante Werte aufweisen würden, weshalb hier nicht zuletzt auch als Denkanstoss für zukünftige Forschung nicht-signifikante Tendenzen ausgewiesen werden sollen.

Im Folgenden kommen Verfahren der uni-, bi- und multivariaten deskriptiven Statistik zum Einsatz. Univariate Analysen werden verwendet, wenn ein einzelnes Merkmal untersucht wird, z.B. bei Betrachtung des Mittelwerts und der Streuung des telc-Resultats. Bivariate Verfahren untersuchen den Zusammenhang zwischen zwei Merkmalen, beispielsweise zwischen Erstschriftsprache und Lernerfolg. Multivariate Verfahren werden angewendet, wenn der Einfluss mehrerer Merkmale gleichzeitig untersucht wird, wie dies beispielsweise bei einer linearen Regression der Fall ist (vgl. Weins 2010, S. 65).

6.2.1 Uni- und bivariate Analyse

Wie bereits zu Beginn des Kapitels erläutert, wird im Folgenden auf die Anzahl erfolgreich absolvierter telc-A1-Prüfungen und die dabei erreichten Punkte als abhängige Variable eingegangen. Wie aus dem Säulendiagramm (s. Abb. 16) ersichtlich, hat rund ein Drittel der 57 Teilnehmenden, welche die telc-Prüfung A1 im Dezember 2018 absolviert haben, diese bestanden, d. h. über 60 % der möglichen Punkte erreicht. Davon haben drei Personen 80–89 % der möglichen Punkte (die Note *gut*), vier Teilnehmende 70–79 % (die Note *befriedigend*) und zwölf Lernende 60–69 % (die Note *ausreichend*) erhalten.

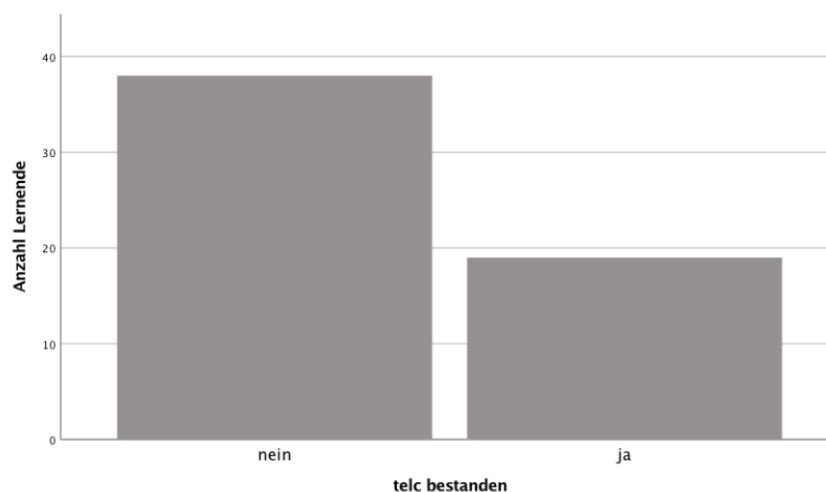


Abb. 16: telc pass/fail

Um zu genaueren Einsichten zu kommen, werden im nachstehenden Säulendiagramm (s. Abb. 17) alle Ergebnisse in 5 %-Schritten dargestellt.

Es ist zu erkennen, dass sich 16 Personen im Bereich 50–59 % befinden und die Prüfung entsprechend eher knapp nicht bestanden haben. Weitere 16 Personen haben zwischen 40 und 49 % der Punkte erreicht und sechs Teilnehmende haben die Prüfung mit unter 40 % eindeutig nicht bestanden. Der Mittelwert der in der Prüfung erreichten Punkte liegt bei 54 % (SD = 13.0). Auch der Median ist unterhalb von 60 % angesiedelt und gibt an, dass die Mehrheit der Teilnehmenden die telc-Prüfung nicht bestanden hat.

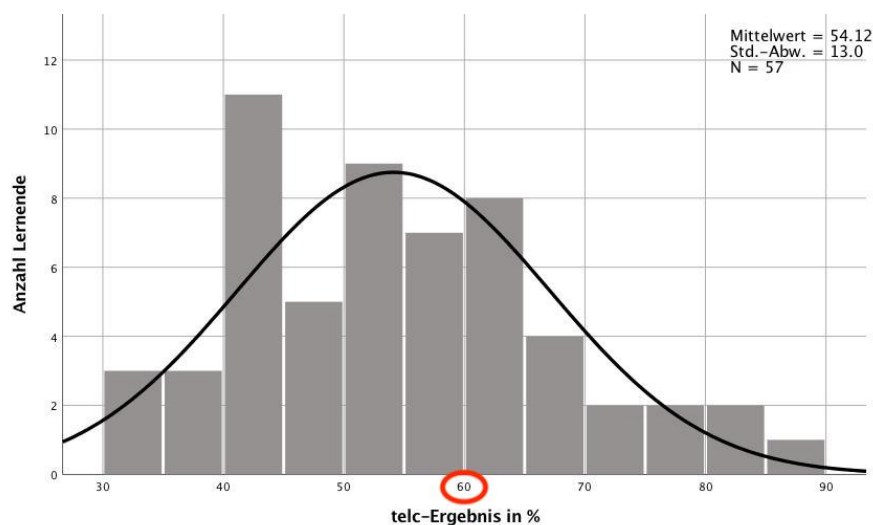


Abb. 17: telc-Prozentwerte

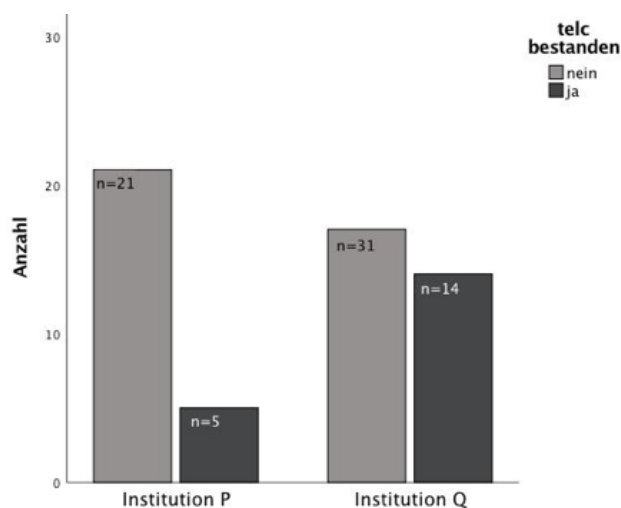


Abb. 18: telc pass/fail differenziert nach Institution

Werden die Ergebnisse in Bezug auf die beiden **Institutionen** unterschieden (s. Abb. 18), so ist festzustellen, dass bei Institution Q ($n = 31$) mit 14 Teilnehmenden (45 %) deutlich mehr Personen die telc-Prüfung bestanden haben als bei Institution P ($n = 26$) mit fünf

Teilnehmenden (19 %). Dieser Zusammenhang ist nicht sehr stark ausgeprägt ($\chi^2(1) = 4.28$, $\Phi = .27$), jedoch statistisch signifikant ($p = .043$).

Ein Vergleich der beiden Institutionen zeigt zudem (s. Abb. 19), dass Teilnehmer der Institution Q im Schnitt besser (57 % der möglichen Punkte) als Teilnehmer von Institution P (50 %) abgeschnitten haben. Nach Durchführung eines t-Tests ($t(55) = -2.07$, $p = .038$) zeigt sich, dass sich die Institutionen bezüglich der Erfolgsquote bei der telc-Prüfung A1 mit einem Signifikanzniveau von .05 und einem mittleren Effekt ($d = .56$) unterscheiden. Festhalten lässt sich somit, dass die höhere Anzahl der erfolgreichen Teilnehmenden bei Institution Q statistisch bedeutsam ist. In Abschnitt 8.2 wird auf diese Unterschiede zwischen den Institutionen vertiefend eingegangen.

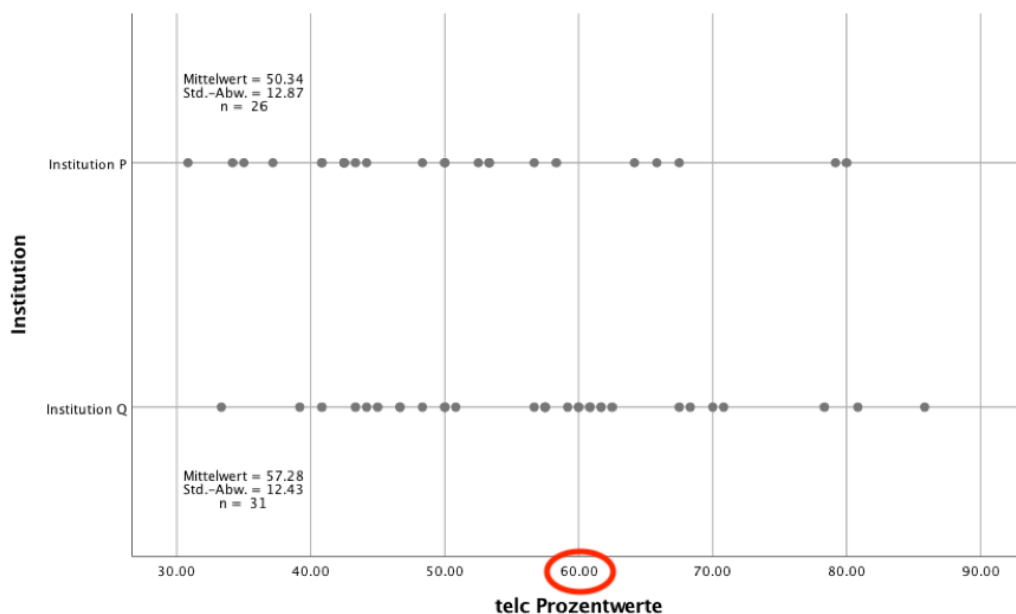


Abb. 19: telc-Prozentwerte differenziert nach Institution

Bei der Betrachtung der **Kurslevel** zeigt sich, dass drei von 19 Teilnehmenden (16 %) im tiefen, sechs von 17 Personen (35 %) im mittleren und zehn von 21 Teilnehmenden (48 %) im hohen Kurslevel die telc-Prüfung erfolgreich absolviert haben. Der Zusammenhang zwischen bestandener telc-Prüfung und Kurslevel ist statistisch nicht signifikant ($\chi^2(2) = 4.59$, $p = .101$). Aus der unteren Boxplot-Grafik (s. Abb. 20) geht hervor, dass Teilnehmer im tiefen Kurslevel 47 %, im mittleren 53 % und im hohen 61 % der Punkte in der telc-Prüfung erreicht haben. Die hohen Kurslevel weisen den höchsten Mittelwert auf, was mit den entsprechend höheren Vorkenntnissen bei Kursstart erklärt werden kann. Nach Durchführung einer einfaktoriellen Varianzanalyse (ANOVA) zeigt sich, dass das Kurslevel einen signifikanten Einfluss auf das telc-Ergebnis in Prozent hat ($F(2,54) = 6.8$, $p = .002$, partielles $\eta^2 = .20$). 17 % (korrigiertes R^2) der

Streuung der telc-Prozentwerte können durch das Kurslevel erklärt werden. Das $\eta^2 = .20$ entspricht einer starken Effektstärke. Post-hoc-Tests mit Bonferroni-Korrektur zeigen, dass sich die beiden Pole *Kurslevel tief* ($M = 47.48$, $SD = 10.32$) und *Kurslevel hoch* ($M = 61.15$, $SD = 13.04$) signifikant ($p = .002$) unterscheiden. Das mittlere Kurslevel ($M = 52.84$, $SD = 11.83$) unterscheidet sich hingegen nicht signifikant von den beiden anderen Kursniveaus. Das höchste Level scheint am effektivsten in Bezug auf den Prüfungserfolg zu sein.

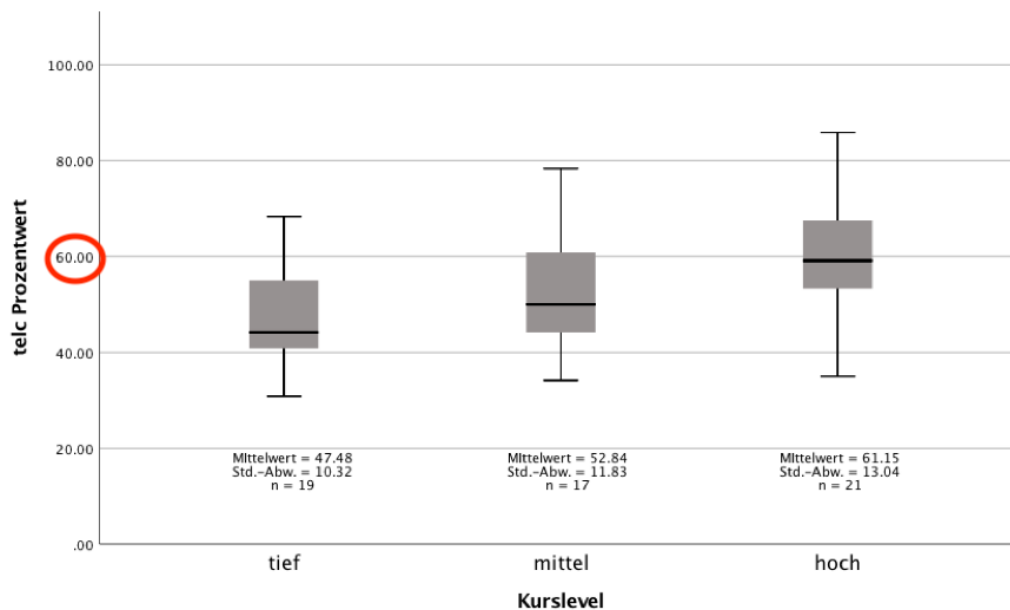


Abb. 20: telc-Prozentwerte differenziert nach Kurslevel

Die relative **Geschlechterverteilung** ist weitgehend ausgeglichen; elf männliche (31 %, $n = 35$) und acht weibliche (36 %, $n = 22$) Teilnehmende haben die telc-Prüfung bestanden. Statistisch ist die Chance, die telc-Prüfung zu bestehen, nicht signifikant an das Geschlecht gekoppelt ($\chi^2(1) = .15$, $p = .700$). Werden die in den telc-Prüfungen erreichten Punkte der weiblichen und männlichen Teilnehmer verglichen, so fällt auf, dass weibliche Teilnehmer einen Durchschnittswert von $M = 53$ % der erworbenen Punkte erzielten, wohingegen männliche Teilnehmer durchschnittlich 55 % erlangten. Die Ergebnisse des t-Tests ($t(55) = .65$, $p = .516$) lassen darauf schließen, dass die bei der telc-Prüfung A1 erreichten Punkte ebenfalls nicht statistisch signifikant durch das Geschlecht beeinflusst werden.

Die Variable **Alter** wird anhand von Clustern analysiert³³: bis 19 Jahre, 20 bis 29 Jahre, 30 bis 39 Jahre, 40 bis 49 Jahre und über 50, wobei zwei Personen 51 und ein Teilnehmer 19 Jahre

³³ Ein bekanntes Problem bei der Altersangabe von Flüchtlingen ist, dass Geburtsdaten oft unbekannt, ungenau oder falsch sind. Deshalb wird bei der Einreise in die Schweiz oft der 1. Januar des vermuteten Geburtsjahrs als Geburtstag festgelegt. Im Fragebogen der vorliegenden Studie wurde deshalb nur nach dem Geburtsjahr und nicht nach dem genauen Datum gefragt. Der Verständlichkeit und Lesefreundlichkeit halber wird im Fliesstext

alt waren. Die Mehrheit der Teilnehmenden, welche die telc-Prüfung erfolgreich absolviert haben, ist zwischen 30 und 39 Jahre alt (44 %, $n = 16$). Dieses Alterscluster hat auch die im Schnitt höchste Prozentpunktzahl erreicht ($M = 56$ %, $SD = 13.82$), gefolgt von den 40- bis 49-jährigen Lernenden ($M = 40$ %, $n = 10$) mit 53 der 100 möglichen Punkte ($SD = 13.03$). Von den Teilnehmenden zwischen 20 und 29 Jahren haben acht Personen die Prüfung bestanden, was 29 % der gesamten Altersgruppe entspricht, bei einer durchschnittlichen Prozentpunktzahl von 55 % ($SD = 12.94$, $n = 28$). Die drei Personen, die 19 bzw. 51 Jahre alt waren, haben die Prüfung alle nicht bestanden. Nach Durchführung einer einfaktoriellen Varianzanalyse zeigt sich, dass die Varianz der unabhängigen Variable zwischen den Alterscluster nicht statistisch signifikant ist ($F(3,52) = .80$, $p = .534$, partielles $\eta^2 = .06$). Ein Zusammenhang zwischen Alter und erfolgreich bestandener telc-Prüfung lässt sich statistisch ebenfalls nicht begründen ($\chi^2(4) = 2.77$, $p = .598$).

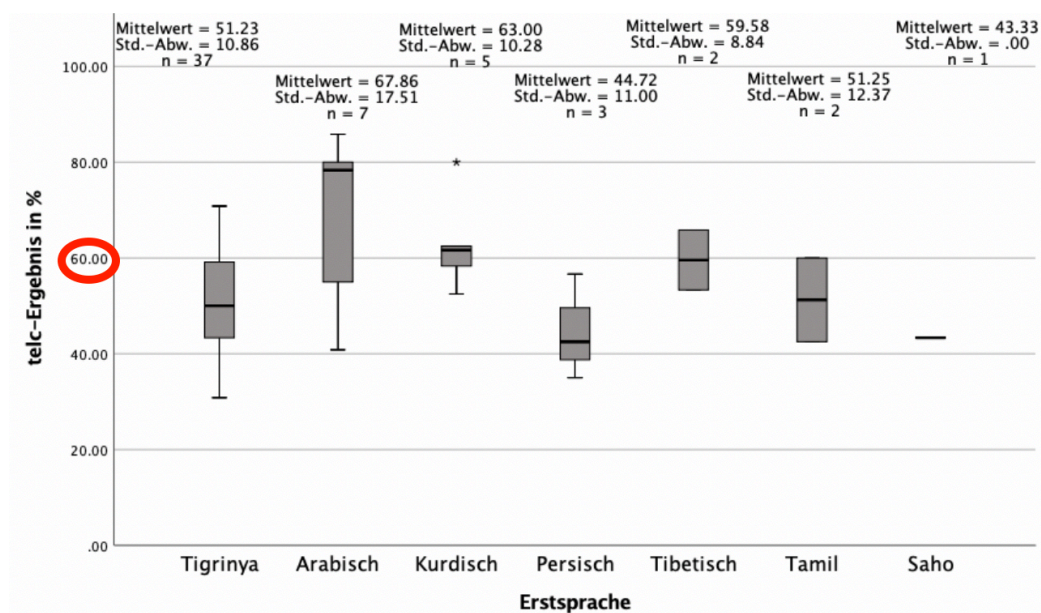


Abb. 21: telc-Prozentwerte differenziert nach L1

Lernende mit **Erstsprache** Persisch ($M = 45$ %, $SD = 11.00$) zeigen den geringsten Mittelwert und haben die telc-Prüfung alle nicht bestanden (s. Abb. 21). Nur knapp darüber liegt der Einzelfall mit Saho als Erstsprache ($M = 43$ %, $SD = .0$). Bei Tamil ($M = 51$ %, $SD = 12.37$) und Tibetisch ($M = 60$ %, $SD = 8.84$) hat jeweils die Hälfte die Prüfung bestanden, was je einer Person entspricht. Insgesamt neun von 37 Teilnehmern mit Tigrinya als Erstsprache ($M = 51$ %, $SD = 1.86$) haben die Prüfung erfolgreich absolviert, was einer Erfolgsquote von 24 %

aber das Alter und nicht das Geburtsjahr angegeben. Um das Alter zu berechnen, wurde diese Variable so konstruiert, dass alle Teilnehmenden im Jahr 2018 ein Jahr älter wurden, d. h. der Stichtag war der 31.12.2018.

entspricht. Bei Arabisch als Erstsprache hat fast drei Viertel die Prüfung bestanden ($M = 71 \%$, $n = 7$) und überdies die höchsten Prozentpunkte erreicht ($M = 68 \%$, $SD = 17.51$). Auch Personen mit Kurdisch als Erstsprache waren relativ erfolgreich: Drei der insgesamt fünf Teilnehmer haben die Prüfung erfolgreich abgeschlossen und mit 63% ($SD = 1.28$) der möglichen Punkte weisen sie den zweithöchsten Mittelwert auf.

Der Zusammenhang zwischen Erstsprache und der Wahrscheinlichkeit, die telc-Prüfung A1 zu bestehen, ist statistisch nicht signifikant ($\chi^2(6) = 1.02$, $p = .124$). Es existiert jedoch ein signifikanter Zusammenhang zwischen der Erstsprache und den erreichten Punkten bei der telc-Prüfung ($F(5,50) = 2.96$, $p = .015$, partielles $\eta^2 = .26$), mit einem starken Effekt.

Bei einer exakteren Betrachtung der Erstschriftsprache, die über die Beurteilung des **Wortschatzes** erfolgt, (s. Abschn. 4.3.2) fällt auf, dass Teilnehmende mit einer Bewertung auf C1/C2-Niveau durchschnittlich den höchsten Prozentwert, 55% der in der der telc-Prüfung erreichten Punkte ($SD = 12.0$), aufweisen und diese dementsprechend auch zahlenmässig am häufigsten bestanden haben (14 von $n = 37$). Lernende mit einer Bewertung auf B1/B2-Niveau erreichen den zweithöchsten Prozentpunktwert und lediglich einen geringfügig niedrigeren als Personen im C-Niveau (55% , $SD = 18.0$). Insgesamt drei Personen mit Wortschatzkenntnissen im B-Bereich der Erstschriftsprache haben die telc-Prüfung bestanden ($n = 11$). Teilnehmende mit einer Bewertung auf A2-Niveau, dem am tiefsten gemessenen Niveau in der Erstschriftsprache, erzielten den geringsten durchschnittlichen telc-Prozentpunktwert mit 50% ($SD = 1.16$), wobei zwei Personen die Prüfung erfolgreich absolviert haben ($n = 9$). Um zu sehen, ob sich diese Mittelwerte statistisch signifikant unterscheiden, wurde auch hier eine ANOVA durchgeführt. Die Ergebnisse dieser Analyse zeigen, dass diese Unterschiede statistisch nicht signifikant sind ($F(2,54) = .64$, $p = .532$, partielles $\eta^2 = .02$). Das gilt auch für die pass/fail-Quote ($\chi^2(2) = 1.02$, $p = .061$).

Ein ähnliches Ergebnis ergibt sich hinsichtlich der **formalen Richtigkeit** in der Erstschriftsprache: Mit der höchsten Bewertung auf C-Niveau haben sechs Personen ($n = 17$) die telc-Prüfung bestanden, im B-Bereich sieben Lernende ($n = 21$) und im A2-Bereich sechs Teilnehmende ($n = 19$). Auch hier erreichten Personen mit der niedrigsten Bewertung auf A2-Niveau die geringste Punktzahl in der telc-Prüfung (53% , $SD = 16.49$). Die meisten telc-Prozentpunkte erzielten Teilnehmende mit einer Bewertung auf B1/B2-Niveau (55% der erreichten Punkte, $SD = 10.39$), gefolgt von Lernenden auf C-Niveau (55% , $SD = 12.18$). Bei genauerer Betrachtung der formalen Richtigkeit fällt auf, dass gleich viele Personen mit einer

Bewertung auf A2-Niveau die telc-Prüfung bestanden haben wie solche mit einer Bewertung auf C-Niveau ($n = 6$). Sieben Teilnehmende, und somit die grösste Gruppe der erfolgreichen Absolventen der telc-Prüfung A1, hatten eine Bewertung auf B-Niveau. Ebenfalls zeigt sich hier nach Durchführung einer ANOVA, dass zwischen formaler Richtigkeit in der Erstschriftsprache und erreichten Punkten in der Prüfung kein statistisch bedeutsamer Zusammenhang besteht ($F(2,54) = .09, p = .910$, partielles $\eta^2 = .00$). Das gilt auch für die Erfolgsquote bei der Prüfung ($\chi^2(2) = .06, p = .973$).

Die von den Teilnehmenden beherrschten **Fremdsprachen** wurden zu Sprachgruppen zusammengestellt: Roman-alphabet, falls der Teilnehmer ein lateinisches Alphabet beherrscht (z. B. Englisch), non-Roman-alphabet bei einem nichtlateinischen Schriftsystem (z. B. Amharisch) und non-alphabet bei einer logografischen Fremdsprache (z. B. Chinesisch). Von insgesamt 13 Teilnehmenden, die eine Fremdsprache beherrschten, haben fünf Personen die telc-Prüfung bestanden (39 %, $n = 5$). Ein Mittelwertvergleich zwischen den Personen mit und ohne Fremdsprachenkenntnisse zeigt, dass Erstere mit 58 % ($SD = 1.08, n = 5$) gegenüber Letzteren mit 53 % ($SD = 13.61, n = 44$) der möglichen Punkte einen höheren Mittelwert aufweisen. Etwas weniger ausgeprägt ist der Unterschied, wenn die Fremdsprache auf einem non-Roman-alphabet ($M = 60 \%, SD = 11.55, n = 7$) statt einem Roman-alphabet basiert ($M = 57 \%, SD = 8.48, n = 6$). Im Falle der Kenntnisse eines non-alphabet-Schriftsystems, Chinesisch, handelt es sich um einen Einzelfall ($M = 53 \%, SD = .00$). Eine einfaktorielle Varianzanalyse zeigt, dass Fremdsprachen keinen signifikanten Einfluss auf die telc-Prozentwerte haben ($F(1,55) = 1.63, p = .207$, partielles $\eta^2 = .03$), ebenso stehen sie nicht im Zusammenhang mit dem Bestehen der telc-Prüfung ($\chi^2(1) = .20, p = .655$).

In Bezug auf die **Anzahl der Schuljahre** wurden folgende Cluster gebildet: 1 bis 3 Schuljahre ($n = 5$), 4 bis 6 Schuljahre ($n = 13$), 7 bis 9 Schuljahre ($n = 29$) und 10 bis 12 Schuljahre ($n = 9$). Lediglich eine Person gab an, keine Schule im Heimatland besucht zu haben. Die Anzahl der Schuljahre hat einen positiven Einfluss auf die Wahrscheinlichkeit, die telc-Prüfung zu bestehen. Während das Individuum ohne Schulbildung die erforderliche Punktzahl bei der Abschlussprüfung nicht erreicht hat, hat eine von fünf Personen mit 1 bis 3 Jahren Schulerfahrung die Prüfung bestanden. Die Erfolgsquote der nachfolgenden Cluster liegt bei 31 % respektive 35 %. Die Teilnehmenden mit den meisten Schuljahren, d. h. mit 9–12 Jahren, haben die Prüfung in 44 % aller Fälle bestanden. Die Person ohne Schulerfahrung hat 43 % der möglichen Punkte erzielt, die Gruppen der Individuen mit 1 bis 3 und 4 bis 6 Schuljahren

weisen einen Mittelwert von je rund 52 % ($SD = 8.76$, $SD = 18.08$) auf. Lernende mit 7 bis 9 Schuljahren weisen einen Mittelwert von 55 % ($SD = 1.96$) und Teilnehmende mit 10 bis 12 Schuljahren einen Wert von 58 % ($SD = 13.78$) der erreichbaren Punkte auf. Die Resultate deuten darauf hin, dass sich die Schulerfahrung positiv auf das Bestehen der telc-Prüfung auswirkt; je höher die Anzahl der Schuljahre, desto bessere Ergebnisse erzielten die Teilnehmenden. Es zeigt sich in der ANOVA jedoch kein statistisch signifikanter Einfluss der Schuljahre auf die telc-Prozentwerte ($F(3,52) = .50$, $p = .736$, partielles $\eta^2 = .04$), ebenso wenig lässt sich ein Zusammenhang zwischen der Anzahl Schuljahre und erfolgreichem Abschneiden in der telc-Prüfung nachweisen ($\chi^2(4) = 1.46$, $p = .834$).

Einer **Arbeit** im Heimatland sind 35 Personen nachgegangen, wovon neun Personen eine **Berufslehre** oder eine ähnliche Ausbildung abgeschlossen haben. Von den zuvor berufstätigen Teilnehmenden haben zwölf Personen (34 %, $n = 35$) die telc-Prüfung erfolgreich absolviert, ausserdem drei Personen (33 %, $n = 9$) unter denen, die über eine Berufsausbildung verfügen. Die Arbeitserfahrung hat in der vorliegenden Stichprobe keinen Einfluss auf die erreichten Punkte bei der telc-Prüfung, Teilnehmende mit und ohne Arbeitserfahrung haben im Schnitt 54 der möglichen 100 Punkte erreicht ($SD = 13.88$, $SD = 11.76$). Dasselbe gilt für Berufslehren, auch hier haben Teilnehmende mit und ohne Berufslehre gleich viele Punkte erreicht ($M = 54$ %, $SD = 9.86$, $SD = 13.59$). Nach Durchführung einer ANOVA zeigt weder die Variable Arbeit ($t(55) = .13$, $p = .900$) noch Berufslehre ($t(55) = 1.82$, $p = .973$) statistische Signifikanz. Ebenso ergeben die Variablen Arbeit ($\chi^2(1) = .04$, $p = .847$) und Berufslehre ($\chi^2(1) = .00$, $p = 1.000$) keinen Zusammenhang mit dem erfolgreichen Abschneiden bei der telc-Prüfung.

Hinsichtlich des **Gebrauchs des Deutschen** fällt auf, dass mehr Lernende die telc-Prüfung bestanden haben, die Deutsch vier- bis siebenmal pro Woche ausserhalb des Sprachkurses benutzten (38 %, $n = 24$), als Teilnehmende, die ein- bis dreimal pro Woche auf Deutsch kommunizierten (29 %, $n = 24$). Neun Personen gaben an, dass sie nie Deutsch ausserhalb des Unterrichts benutzten, drei davon haben die telc-Prüfung bestanden (33 %, $n = 9$). Die Mittelwerte der Gruppen unterscheiden sich nur geringfügig: So haben Teilnehmende, die ein- bis dreimal pro Woche Deutsch benutzten, einen Mittelwert von 54 % der erreichten Punkte ($SD = 12.56$, $n = 24$) im Gegensatz zu Teilnehmern, die nie ausserhalb des Unterrichts auf Deutsch kommunizierten mit 51 % ($SD = 1.75$, $n = 9$). Teilnehmer, die Deutsch vier- bis siebenmal pro Woche anwandten, weisen mit $M = 55$ % ($SD = 14.46$, $n = 24$) den höchsten Wert auf. In Bezug auf den Gebrauch des Deutschen zeigen ANOVA ($F(2,54) = .28$, $p = .756$,

partiell $\eta^2 = .01$) und Chi-Quadrat ($\chi^2(2) = .38, p = .829$), dass die Unterschiede nicht signifikant sind.

Insgesamt 46 der 57 untersuchten Personen (81 %) haben bereits vor Teilnahme am Forschungsprojekt einen **Deutschkurs** besucht. Für die restlichen elf Lernenden (19 %) ist der untersuchte Zweitschriftlernerkurs der erste Deutschkurs (s. Abb. 22). Von diesen Teilnehmenden haben sieben Personen (64 %, $n = 11$) die telc-Prüfung bestanden – dies in deutlichem Unterschied zu zwölf Personen (26 %, $n = 46$), die bereits mindestens einen Deutschkurs besucht haben. Wie der unten stehenden Grafik zu entnehmen ist, ist die Durchfallquote bei Teilnehmern mit vorgängiger Deutschkursteilnahme deutlich höher als bei Personen, die zum ersten Mal einen Deutschkurs besuchten. Dieser negative Zusammenhang zwischen vorgängig besuchtem Deutschkurs und der Wahrscheinlichkeit, die A1-Prüfung zu bestehen, ist signifikant ($\chi^2(1) = 5.63, p = .022$) und mittelstark ausgeprägt ($\Phi = -.31, p = .022$). Der Mittelwert der erreichten Punkte von Personen, die vorgängig noch keinen Deutschkurs besucht haben, ist ebenfalls deutlich höher (62 % der erreichten Punkte, $SD = 16.47$) als bei den Personen, die bereits mind. einen Deutschkurs besucht haben (52 %, $SD = 11.43$). Somit schneiden Lernende, die vorgängig noch keinen Deutschkurs besucht haben, besser ab und dies statistisch signifikant ($t(55) = 2.37, p = .022$). Die Effektstärke $d = .65$ entspricht einem mittleren Effekt.

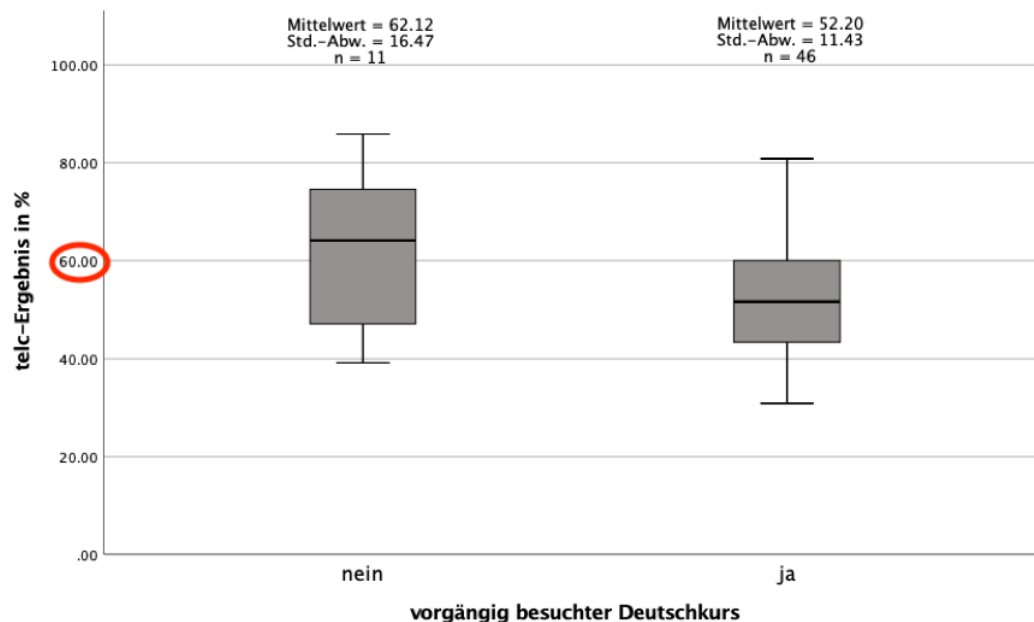


Abb. 22: telc-Prozentwerte differenziert nach vorgängig besuchtem Deutschkurs

Von den insgesamt 19 Teilnehmenden, welche die Prüfung bestanden haben, wurden zehn Lernende in Kursen mit zusätzlichem Fokus auf der **Silbenmethode** unterrichtet (33 %, $n = 30$,

Treatmentgruppen 2). Neun Teilnehmende aus den Treatmentgruppen 1, d. h. ohne Silbenmethode, haben die telc-Prüfung bestanden (35 %, $n = 26$), die Erfolgsquote ist bei beiden Treatments etwa gleich hoch. Aus einem Vergleich der Kurslevel wird jedoch ersichtlich, dass die einzigen Teilnehmenden, welche die telc-Prüfung aus dem tiefsten Kurslevel bestanden haben, mit der Silbenmethode unterrichtet wurden. Im mittleren Kurslevel ($n = 9$) haben 40 % der Teilnehmenden mit zusätzlichem Treatment die Prüfung bestanden, im Vergleich zu 28 % ohne Silbenmethode ($n = 11$). Im höchsten Kurslevel hingegen ist zu sehen, dass die Erfolgsquote bei der telc-Prüfung für die Teilnehmenden, die ohne Silbenmethode unterrichtet wurden, doppelt so hoch ist (64 %, $n = 11$) wie für diejenigen, die mit der Silbenmethode unterrichtet wurden (30 %, $n = 10$). Allerdings ist der Zusammenhang zwischen Kurslevel und erfolgreich bestandener telc-Prüfung statistisch nicht signifikant ($\chi^2(5) = 9.527, p = .090$), (Cramer $V = .409, p = .090$).

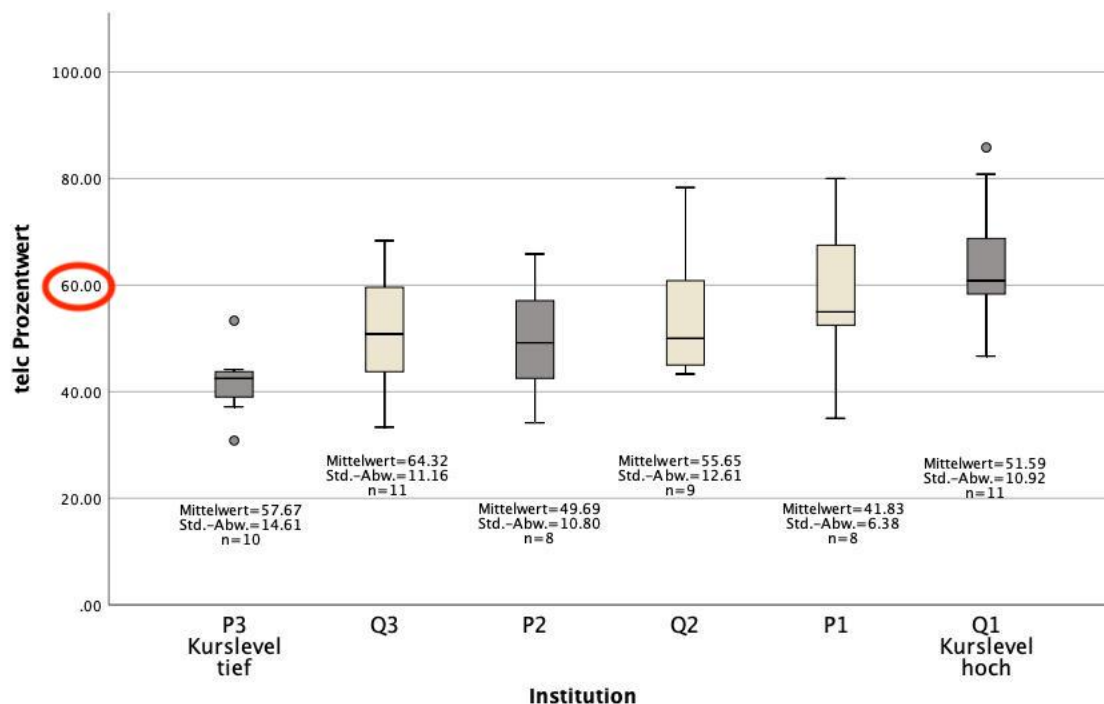


Abb. 23: telc-Prozentwerte differenziert nach Kurslevel und Silbenmethode
hell: mit Silbenmethode, dunkel: ohne Silbenmethode

Ein Mittelwertvergleich der erreichten Punkte in Kursen mit und ohne Silbenmethode weist nur einen geringfügigen Unterschied auf (s. Tab. 40): So zeigen Kurse mit der Silbenmethode einen Mittelwert von $M = 55$ % der erreichten Punkte ($SD = 12.58$) und die Kurse ohne Silbenmethode einen Wert von $M = 53$ % ($SD = 13.64$). Der Median für die Silbenmethode liegt bei 53 % und für Kurse ohne Silbenmethode bei 50 %. Bezüglich Kurslevel schneiden die Kurse mit der Silbenmethode auf dem tiefen und dem mittleren Niveau besser ab, im hohen Kurslevel

hingegen erzielte die Gruppe, die ohne Silbenmethode unterrichtet wurde, bessere Ergebnisse (s. Abb. 23). Nach Durchführung von t-Tests wird jedoch erkennbar, dass lediglich die Varianz auf dem tiefen Kurslevel auf einem .05-Signifikanzniveau bedeutend ist. Der Levene-Test ($F(1,17)=3.646, p = .073$) zeugt von einer (schwachen) Varianzhomogenität der zu vergleichenden Gruppen auf dem tiefen Kurslevel. Der entsprechende t-Test ($t(45)= -2.253, p = .038$) zeigt, dass die durchschnittlich erreichten Punkte der Gruppe mit Silbenmethode ($M = 51.59, SD = 1.82, n = 11$) signifikant über denjenigen der Gruppe ohne Silbenmethode ($M = 41.83, SD = 6.38, n = 8$) liegen. Die dazugehörige Effektgrösse $d = -1.0$ entspricht einem starken Effekt. Auf dem mittleren und dem hohen Kursniveau sowie über alle Niveaus hinweg unterscheiden sich die Mittelwerte hingegen nicht signifikant.

	Silbenmethode	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
telc-Ergebnis in %, alle Kurslevel	ja	30	54.83	12.58	2.30
	nein	27	53.32	13.64	2.63
telc-Ergebnis in %, Kurslevel tief	ja	11	51.59	10.92	3.29
	nein	8	41.83	6.38	2.25
telc-Ergebnis in %, Kurslevel mittel	ja	9	55.65	12.61	4.20
	nein	8	49.69	10.80	3.82
telc-Ergebnis in %, Kurslevel hoch	ja	10	57.67	14.61	4.62
	nein	11	64.32	11.16	3.37

Tab. 40: telc-Prozentwerte differenziert nach Kurslevel und Silbenmethode

6.2.2 Die Treatments im Vergleich

Im Folgenden sollen die Kurse mit und ohne Silbenmethode genauer hinsichtlich des Abschneidens der einzelnen Teilnehmenden bei der A1-Prüfung untersucht werden.

Werden die Gruppen nicht nach Kurslevel differenziert und wird das Abschneiden in den telc-Fertigkeiten über alle drei Kurslevel mit und ohne Silbenmethode betrachtet (s. Tab. 41), so ist ein besseres Abschneiden der Silbenmethode bei allen Fertigkeiten ausser dem Hören erkennbar. Dieser Zusammenhang ist als Trend aus den Daten ablesbar, eine statistische Signifikanz liegt allerdings nach Durchführung von t-Tests nicht vor ($p \geq .05$).

		telc-Ergebnis in Prozenten	telc Hören	telc Lesen	telc Schreiben	telc Sprechen
Kurslevel total ohne Silbenmethode	Mittelwert	53.32	52.93	48.89	37.28	74.81
	n	27	27	27	27	27
	Std.-Abweichung	13.64	16.88	13.46	27.57	18.43
Kurslevel total mit Silbenmethode	Mittelwert	54.83	50.67	50.00	41.56	77.10
	n	30	30	30	30	30
	Std.-Abweichung	12.58	17.91	13.53	23.22	19.51
Kurslevel tief ohne Silbenmethode	Mittelwert	41.83	39.17	51.67	19.58	57.92
	n	8	8	8	8	8
	Std.-Abweichung	6.38	9.04	11.13	13.38	15.22
Kurslevel tief mit Silbenmethode	Mittelwert	51.59	50.91	50.91	37.88	66.64
	n	11	11	11	11	11
	Std.-Abweichung	10.92	10.45	13.42	23.77	21.75
Kurslevel mittel ohne Silbenmethode	Mittelwert	49.69	58.34	45.83	24.17	70.42
	n	8	8	8	8	8
	Std.-Abweichung	10.80	15.84	16.50	21.44	13.85
Kurslevel mittel mit Silbenmethode	Mittelwert	55.65	44.44	46.67	42.96	88.52
	n	9	9	9	9	9
	Std.-Abweichung	12.61	23.57	10.54	27.41	7.47
Kurslevel hoch ohne Silbenmethode	Mittelwert	64.32	58.18	49.09	59.70	90.30
	n	11	11	11	11	11
	Std.-Abweichung	11.16	17.41	13.42	24.24	8.88
Kurslevel hoch mit Silbenmethode	Mittelwert	57.67	56.00	52.00	44.33	78.33
	n	10	10	10	10	10
	Std.-Abweichung	14.61	18.64	16.57	20.37	19.64
Insgesamt	Mittelwert	54.12	51.58	49.47	39.53	76.02
	N	57	57	57	57	57
	Std.-Abweichung	13.00	17.30	13.39	25.23	18.87

Tab. 41: telc-Prozentwerte unterteilt nach Fertigkeit und differenziert nach Kurslevel und Silbenmethode

Diese Ausführungen machen deutlich, dass keine generellen Aussagen über die Wirkung der Silbenmethode auf die Fertigkeiten getätigt werden können, vielmehr ist nach Kurslevel und Fertigkeit zu differenzieren. t-Tests in Bezug auf das Abschneiden der einzelnen Kursgruppen in den jeweiligen Fertigkeiten bei der telc-A1-Prüfung bestätigen dies mehrheitlich. Von allen zwölf paarweisen t-Tests (für jedes der Kurslevel wurden die Varianten mit und ohne Silbenmethode zu den vier Fertigkeiten in Beziehung gesetzt) fallen lediglich zwei statistisch signifikant aus, und zwar jeweils zugunsten der Silbenmethode. Im tiefen Kurslevel (s. Abb. 24) weist die Gruppe mit Fokus auf der Silbenmethode ein deutlich besseres Ergebnis bei der Fertigkeit Hören auf ($t(17) = -2.55, p = .021$), mit einem starken Effekt ($d = -1.24$). Im mittleren Kurslevel (s. Abb. 25) schneidet die Gruppe mit Silbenmethode bei der Fertigkeit Sprechen ($t(15) = -3.411, p = .004$) besser ab, ebenfalls mit einem starken Effekt ($d = -1.76$).

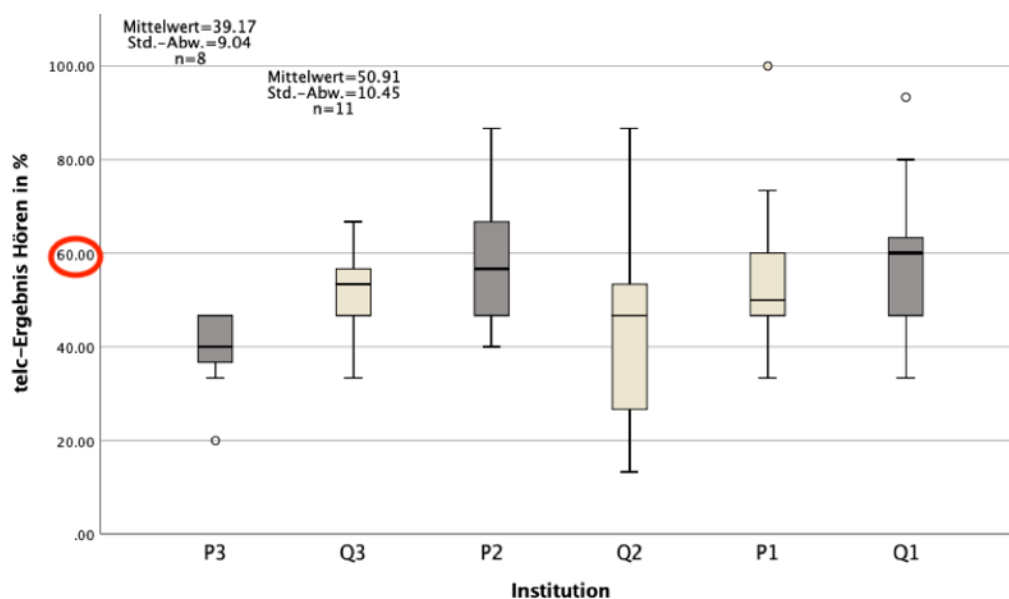


Abb. 24: telc-Prozentwerte Hören differenziert nach Kurslevel und Silbenmethode
hell: mit Silbenmethode, dunkel: ohne Silbenmethode

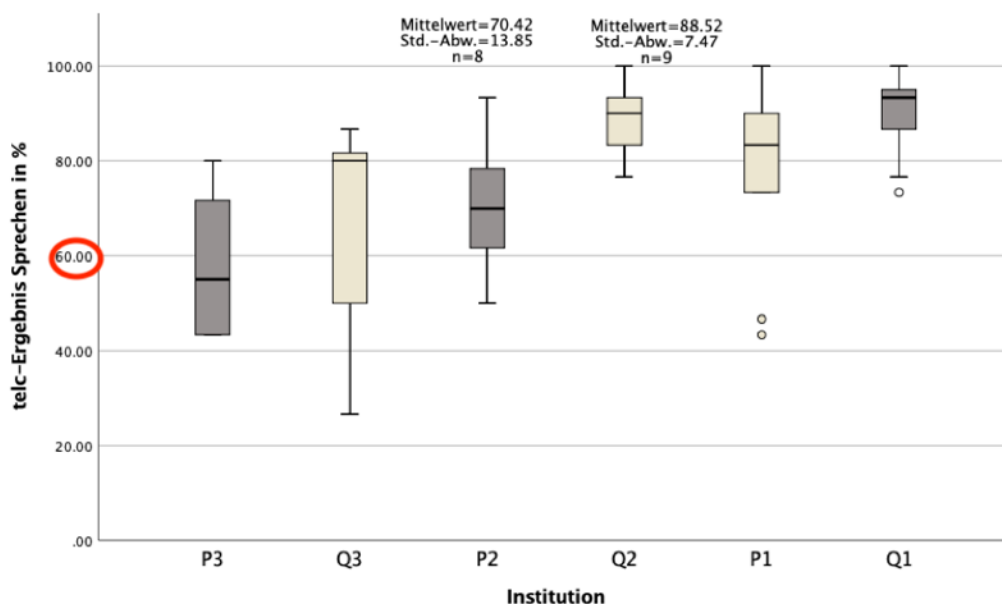


Abb. 25: telc-Prozentwerte Sprechen differenziert nach Kurslevel und Silbenmethode
hell: mit Silbenmethode, dunkel: ohne Silbenmethode

6.2.3 Regression

Bei der linearen Regression³⁴ wurden die Ergebnisse der telc-Prüfung als Regressand und die unabhängige Variablen als Regressoren gewählt. Um Probleme der Multikollinearität zu vermeiden, wurden das *Kurslevel tief*, das *Alterscluster 20 bis 29*, die *L1 Tigrinya, 7 bis 9 Schuljahre* und *kein Deutschgebrauch* ausserhalb des Unterrichts als Referenzkategorien

³⁴ Die Voraussetzungen der linearen Regressionsanalyse wurden überprüft und konnten eingehalten werden: Intervallskala der UV und AV, Linearität des Zusammenhangs, Unabhängigkeit und Normalverteilung des Fehlerwerts sowie die Gauss-Markov-Annahmen 1–5.

bestimmt bzw. aus der Analyse ausgeschlossen. Die Regressionskoeffizienten der anderen Variablen – das Kurslevel, das Alter, die Erstsprache, die Anzahl der Schuljahre und der Gebrauch des Deutschen – verstehen sich folglich immer in Bezug auf diese Regressoren.

Modellzusammenfassung					
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers		
.957 ^a	.916	.795	5.88647		
Koeffizienten ^a					
	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.
	Regressions- koeffizientenB	Std.-Fehler	Beta		
(Konstante)	31.07	8.37		3.72	.001
Institution	10.25	2.45	.40	4.18	.000**
Kurslevel mittel	9.02	2.73	.32	3.31	.003**
Kurslevel hoch	16.60	3.22	.62	5.15	.000**
Silbenmethode	5.02	2.30	.19	2.19	.039*
Geschlecht	7.98	3.04	.30	2.62	.015*
Alter ≤ 19	-2.06	7.81	-.20	-2.57	.017*
Alter 30 bis 39	14.20	2.83	.50	5.03	.000**
Alter 40 bis 49	2.97	3.41	.09	.87	.394
Alter ≥ 50	-3.39	6.03	-.05	-.56	.579
L1 Saho	1.36	9.16	.01	.15	.884
L1 Arabisch	14.66	4.23	.37	3.46	.002**
L1 Kurdisch	6.53	4.22	.14	1.55	.136
L1 Persisch	-17.91	7.61	-.31	-2.35	.028*
L1 Tamil	7.83	7.17	.11	1.09	.286
L1 Tibetisch	3.73	7.03	.05	.53	.600
Wortschatz Erstschrift B1/B2	-8.94	5.65	-.27	-1.58	.127
Wortschatz Erstschrift C1/C2	-6.03	4.85	-.22	-1.24	.226
Formale Richtigkeit Erstschrift B1/B2	1.92	4.38	.07	.44	.666
Formale Richtigkeit Erstschrift C1/C2	2.36	4.32	.08	.55	.591
Fremdsprache	.23	2.95	.01	.08	.940
Schuljahre 0	19.01	9.24	.19	2.06	.051
Schuljahre 1 bis 3	-1.37	4.51	-.03	-.30	.764
Schuljahre 4 bis 6	-8.51	2.86	-.28	-2.98	.007**
Schuljahre 10 bis 12	8.43	4.04	.30	2.09	.048*
Berufslehre	-8.01	3.50	-.23	-2.29	.032*
Universität	15.61	11.10	.16	1.41	.173
Arbeit Heimatland	-6.65	2.33	-.25	-2.86	.009**
Kinder	-1.26	3.11	-.05	-.41	.689
Gesundheitszustand	2.55	1.25	.20	2.04	.054
1–3x pro Woche Deutschgebrauch	16.04	4.78	.62	3.36	.003**
4–7x pro Woche Deutschgebrauch	16.80	4.29	.64	3.92	.001**
vorgängig besuchter Deutschkurs	-21.70	3.50	-.66	-6.20	.000**
Kursabsenz	.06	.12	.05	.47	.640

Tab. 42: Regressionsmodell, AV = telc-Prozentwerte

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

Das Modell weist ein korrigiertes R^2 von .795 ($R^2 = .916$) auf, was bedeutet, dass beinahe 80 % der Varianz der telc-Prozentwerte durch die untersuchten Variablen erklärt werden können. Die unabhängigen Variablen in diesem Modell eignen sich daher, um die abhängige Variable – das telc-Ergebnis – vorherzusagen. Wie aus Tab. 42 ersichtlich, weisen verschiedene Variablen signifikante Koeffizienten auf, die nun *ceteris paribus* – d. h. unter Annahme gleichbleibender anderer Merkmalsausprägungen – betrachtet werden sollen.

Eine erste Variable ist die **Institution**, mit einem hochsignifikanten Wert ($p = .000$). Der Regressionskoeffizient $B = 10.25$ deutet auf ein ausgeprägt höheres telc-Ergebnis der Teilnehmenden aus der Institution Q hin, das um 10.3 Prozentpunkte höher liegt als dasjenige der Institution P.

Das höhere **Kurslevel** weist ebenfalls einen hochsignifikanten Wert ($p = .000$) auf, mit einem Regressionskoeffizienten $B = 16.60$. Demnach haben die höheren Kurslevel durchschnittlich 16 Prozentpunkte mehr erreicht als Kursteilnehmende im tiefen Kurslevel. Das mittlere Kurslevel schneidet ebenfalls besser ab als das tiefste, nämlich um 9 Prozentpunkte ($B = 9.06$, $p = .003$).

Während die **Silbenmethode** sich bei der bivariaten Analyse nur schwach vom Unterricht ohne Silbenmethode abhob, zeugt der Regressionskoeffizient von knapp über 5 ($B = 5.02$, $p = .039$) nun von einem wahrnehmbaren positiven Einfluss dieser Methode auf das telc-Resultat, wenn die anderen Faktoren berücksichtigt und konstant gehalten werden.

Weibliche Teilnehmende weisen durchschnittlich einen um fast 8 Prozentpunkte höheren Wert auf als männliche Lernende ($B = 7.98$, $p = .015$). Teilnehmende zwischen **30 und 39 Jahren** schlossen bei der telc-Prüfung um durchschnittlich 14 Prozentpunkten besser ($B = 14.20$, $p = .000$) ab als 20- bis 29-jährige. Da die Altersgruppe bis 19 Jahre nur einen einzigen Teilnehmer umfasste, ist trotz signifikantem Koeffizient das schlechtere Abschneiden unter Vorbehalt zu betrachten ($B = -2.06$, $p = .017$).

Teilnehmende mit Arabisch als **Erstsprache** haben einen höheren Prozentwert, Lernende mit Persisch als Erstsprache hingegen einen geringeren Wert erreicht als Personen mit Tigrinya als Erstsprache ($B = 14.66$, $p = .002$ resp. $B = -17.91$, $p = .028$).

Teilnehmende, die über 4 bis 6 Jahre **Schulerfahrung** verfügen, erzielten niedrigere telc-Prozentwerte als Teilnehmende mit 7 bis 9 Jahren Schulerfahrung ($B = -8.51$, $p = .007$), und solche mit 10 bis 12 Jahren Schulerfahrung haben höhere telc-Prozentwerte erreicht als die Referenzgruppe ($B = 8.43$, $p = .048$). Zwischen der Dauer der Schulerfahrung und der in der

Prüfung erreichten Anzahl Punkte scheint es demzufolge einen positiven Zusammenhang zu geben.

Der Regressionskoeffizient für Teilnehmer, die im Heimatland einer **Arbeit** nachgegangen sind, ist negativ und statistisch signifikant ($B = -6.65, p = .009$). Ein ähnliches Ergebnis zeigt sich bei der Variable **Berufslehre**: Teilnehmende mit einer Berufslehre im Heimatland weisen einen niedrigeren telc-Prozentwert auf als Lernende ohne Berufslehre ($B = -8.01, p = .032$).

Die Koeffizienten für die **Verwendung des Deutschen** im Alltag liegen ebenfalls im statistisch signifikanten Bereich: Bei ein- bis dreimaligem Gebrauch des Deutschen pro Woche werden bei der telc-Prüfung 16 Punkte mehr erreicht ($B = 16.04, p = .003$), und wer vier bis siebenmal wöchentlich auf Deutsch kommuniziert, erreichte beim Abschlusstest unter ansonsten gleichbleibenden Bedingungen knapp über 16.5 Prozentpunkte mehr ($B = 16.80, p = .001$).

Der Koeffizient der Variable **vorgängig besuchter Deutschkurs** ist hochgradig signifikant ($B = -21.68, p = .000$), was auf einen ausgeprägt geringen durchschnittlichen telc-Prozentwert derjenigen Teilnehmenden hindeutet, die bereits einen Deutschkurs besucht haben: Die Differenz beträgt 21.6 Prozentpunkte.

Die p -Werte für die Variablen keine Schulerfahrung (Schuljahre 0) und Gesundheitszustand, liegen beide knapp über dem .05-Signifikanzniveau.

6.2.4 Überprüfung der Hypothesen

Auf Basis der Befunde zum Einfluss verschiedener Variablen auf das telc-Resultat können nun die Hypothesen H1, H2, H4 und H5 einer Prüfung unterzogen werden.

H1. Mindestens die Hälfte der Kursteilnehmenden der Treatmentgruppen 1 und 2 erreichen das Niveau A1 nach telc innerhalb von 300 Kurslektionen.

Wie aus der deskriptiven Analyse in Abschnitt 6.2.1 ersichtlich ist, erreichten 19 von 57 Teilnehmenden die benötigte Punktzahl in der telc-A1 Prüfung. Das entspricht 33.3 % der Testpersonen. Somit muss H1 falsifiziert werden. Es darf aber angemerkt werden, dass das Erreichen des Niveaus A1 innerhalb von 300 Kurslektionen dennoch möglich ist, wenn auch nur für ein Drittel der getesteten Personen. Werden die beiden Kursinstitutionen gesondert betrachtet, so waren bei einer Institution 45 % der Teilnehmenden bei der telc-Prüfung A1 erfolgreich, was dem Schwellenwert von 50 % relativ nahekommt.

H2. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im telc-Resultat.

H2.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.

H2.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.

- H2.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen
- H2.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.

Die Prüfung der Hypothese H2 stützt sich auf die bivariate Analyse in Abschnitt 6.2.1 und die Regressionsanalyse in 6.2.3. Die Unterhypothesen H2.1–H2.4 werden basierend auf den Ausführungen in Abschnitt 6.2.2 geprüft.

Die Hypothese H2, wonach die Silbenmethode einen Einfluss auf das telc-Ergebnis hat, wird durch die erhobenen Daten verifiziert. Werden die Kurse der Treatmentgruppe 1, d. h. die Kurse ohne Silbenmethode, mit den Kursen der Treatmentgruppe 2, d. h. den Kursen mit Silbenmethode, mit Hilfe eines t-Tests verglichen, ist kein statistisch signifikanter Unterschied feststellbar (s. Abschn. 6.2.2). Der Faktor Silbenmethode weist jedoch in der Regressionsanalyse, unter Konstanthaltung der anderen Variablen, einen signifikanten Regressionskoeffizienten auf, wenn auch einen nicht sehr hohen ($B = 5.02$) (s. Abschn. 6.2.3). Werden zudem die Unterschiede im telc-Resultat zwischen den Treatments je Kurslevel betrachtet, so lässt sich die Varianz im Prüfungsergebnis im tiefen Kurslevel mit der Anwendung der Silbenmethode erklären (s. Abschn. 6.2.2). Festhalten lässt sich damit, dass sich die Silbenmethode schwach positiv auf das telc-Ergebnis auszuwirken und insbesondere im tiefen Kurslevel einflussreich zu sein scheint.

In Bezug auf den Einfluss der Silbenmethode auf die vier Fertigkeiten (H2.1–H2.4) zeigt sich, dass im tiefen Kurslevel mit zusätzlichem Treatment der Silbenmethode in allen Fertigkeiten ausser dem Lesen bessere Ergebnisse erreicht wurden. Im mittleren Kurslevel zeigt die Gruppe mit zusätzlichem Silbenmethoden-Treatment abgesehen von der Fertigkeit Hören ebenfalls überall höhere Werte. Im hohen Kurslevel hingegen konnte für die Treatmentgruppe mit Silbenmethode einzig in der Fertigkeit Lesen ein höheres Ergebnis festgestellt werden. Inferenzstatistisch betrachtet lassen sich jedoch bei zusätzlichem Treatment mit der Silbenmethode nur in der Fertigkeit Hören im tiefen Kurslevel und in der Fertigkeit Sprechen im mittleren Kurslevel signifikant bessere Ergebnisse nachweisen. Im kurslevelübergreifenden Vergleich lassen sich bei keiner Fertigkeit statistisch signifikante Unterschiede feststellen. Die Unterhypothesen H2.1 bis H2.4 müssen deshalb alle als falsifiziert betrachtet werden, wobei bei genauer Betrachtung von H2.3 und H2.4 erkennbar wird, dass zumindest für die Fertigkeiten Hören und Sprechen je nach Kursniveau die Kurse mit Silbenmethode besser abschneiden.

H4. Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat und den Gesamtlernerfolg.

H4.1 Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat

Wie der bivariaten Analyse (s. Abschn. 6.2.1) zu entnehmen ist, hat die Anzahl der Schuljahre einen positiven Einfluss auf die Wahrscheinlichkeit, die telc-Prüfung A1 zu bestehen. Die Erfolgsquote steigt mit jedem Schuljahre-Cluster an, auch wenn der Zusammenhang statistisch nicht signifikant ist. Aus der Regressionsanalyse (s. Abschn. 6.2.3) geht allerdings hervor, dass 4 bis 6 Schuljahre einen signifikant niedrigeren Regressionskoeffizienten ($B = -8.51$) aufweisen im Vergleich zur Referenzkategorie mit 7 bis 9 Schuljahren, das Cluster mit 10 bis 12 Schuljahren hingegen einen signifikant höheren Koeffizienten ($B = 8.43$). Die Hypothese H4.1 kann durch das Regressionsmodell verifiziert werden; eine höhere Anzahl an Schuljahren führt zu höheren telc-Prozentpunkten.

H5. Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das Resultat im telc-Test und den Gesamtlernerfolg.

H5.1 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das telc-Resultat.

Obwohl in der bivariaten Analyse (s. Abschn. 6.2.1) ein Unterschied zwischen den Mittelwerten und dem Wortschatz bzw. der formalen Richtigkeit in der Erstschrift beobachtet werden kann, erweist sich dieser als nicht signifikant. Die Regressionsanalyse (S. Abschn. 6.2.3) weist in die gleiche Richtung, sodass H5.1 falsifiziert werden muss.

6.3 Deskriptive Analyse des Gesamtlernerfolgs bei den forschergenerierten Tests

Damit verschiedene statistische Analysen durchgeführt werden können, müssen die einbezogenen intervallskalierten Variablen normalverteilt sein. Der Shapiro-Wilk-Test zeigt, dass dies der Fall ist ($p = .200$) (s. Abb. 26).

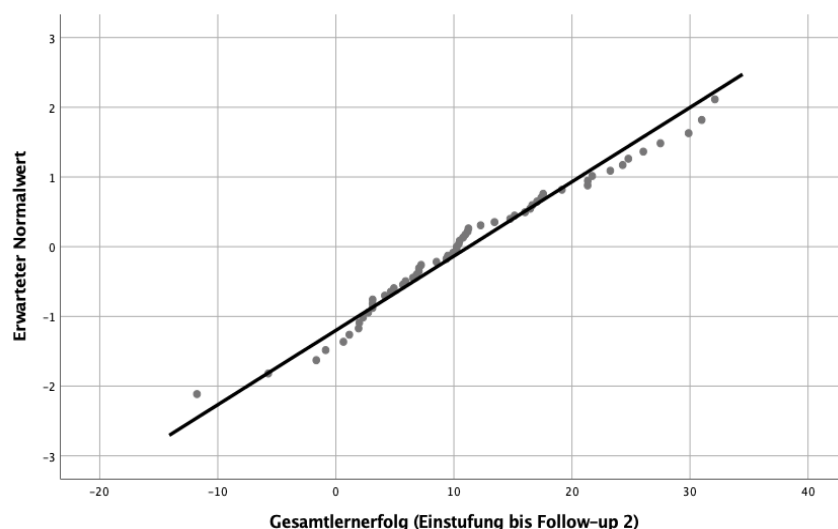


Abb. 26: Gesamtlernerfolg (Punkte) und Normalverteilung (Linie)

Als Übersicht der nachfolgenden statistischen Analysen wird auf unten stehende Tabelle (s. Tab. 43) verwiesen. Wie bereits im Analyseteil zu den telc-Resultaten werden hier alle Variablen diskutiert, unabhängig davon, ob signifikante Werte vorliegen.

Zur Erinnerung, die den folgenden Analysen zugrunde liegende abhängige Variable *Gesamtlernerfolg* berechnet sich als Differenz der durch die Teilnehmenden am Einstufungstest und der zweiten Follow-up-Messung erzielten Punktzahl bei den forschergenerierten Tests Elicited Imitation, Oral Reading Fluency, Hörverstehen und Diktat.

Variable	Deskriptive Analyse	Lineare Regression
Institution	nicht sign.	nicht sign.
Kurslevel	Kurslevel tief ($M = 15$) > Kurslevel hoch ($M = 6$): $F(2,54) = 5.48, p = .007$, partielles $\eta^2 = .17$	nicht sign.
Geschlecht	nicht sign.	nicht sign.
Alter	nicht sign.	Alter ≤ 19 : $B = 24.33$, $p = .012$
Erstsprache (L1)	nicht sign.	nicht sign.
Wortschatzbeherrschung Erstschrift	nicht sign.	nicht sign.
Formale Richtigkeit Erstschrift	nicht sign.	nicht sign.
Fremdsprache	Monolinguale Lernende ($M = 13$) > mehrsprachige Lernende ($M = 5$): $t(55) = 2.94, p = .005, d = .95$ Roman-alphabet ($M = 7$) > non-Roman-alphabet ($M = 6$): $F(1,55) = 8.67, p = .005$, partielles $\eta^2 = .14$	nicht sign.
Anzahl Schuljahre	nicht sign.	nicht sign.
Arbeit Heimatland	nicht sign.	nicht sign.
Berufslehre	Keine Berufslehre ($M = 12$) > Berufslehre ($M = 6$): $t(55) = 2.02, p = .048, d = .75$	nicht sign.
Universität	nicht sign.	nicht sign.
Deutschgebrauch	Kein Deutschgebrauch ($M = 23$) > 4–7x/Woche ($M = 11$) > 1–3x/Woche ($M = 7$): $F(2,54) = 13.66, p = .000$, partielles $\eta^2 = .34$.	nicht sign.
Besuchter Deutschkurs	Kein Deutschkurs ($M = 19$) > besuchter Deutschkurs ($M = 9$): $t(55) = 3.13, p = .003, d = 1.07$	nicht sign.
Silbenmethode	nicht sign.	nicht sign.

Tab. 43: Übersichtstabelle statistische Analysen Gesamtlernerfolg
Punktezuwachs Einstufung bis Follow-up 2

6.3.1 Uni- und bivariate Analyse

Wird die Lernprogression vom Einstufungstest hin zur zweiten Follow-up-Messung anhand der **Institution** differenziert (s. Abb. 27), so weist Institution Q einen durchschnittlichen Lernzuwachs von 11.9 Prozentpunkten ($SD = 8.73, n = 31$) auf und Institution P einen Zuwachs von 10.5 Prozentpunkten ($SD = 10.46, n = 26$). Um zu testen, ob sich diese Mittelwerte signifikant voneinander unterscheiden, wird ein t-Test durchgeführt. Dieser zeigt, dass sich die Institutionen bezüglich Gesamtlernerfolg nicht signifikant unterscheiden ($t(55) = -.59, p = .560$).

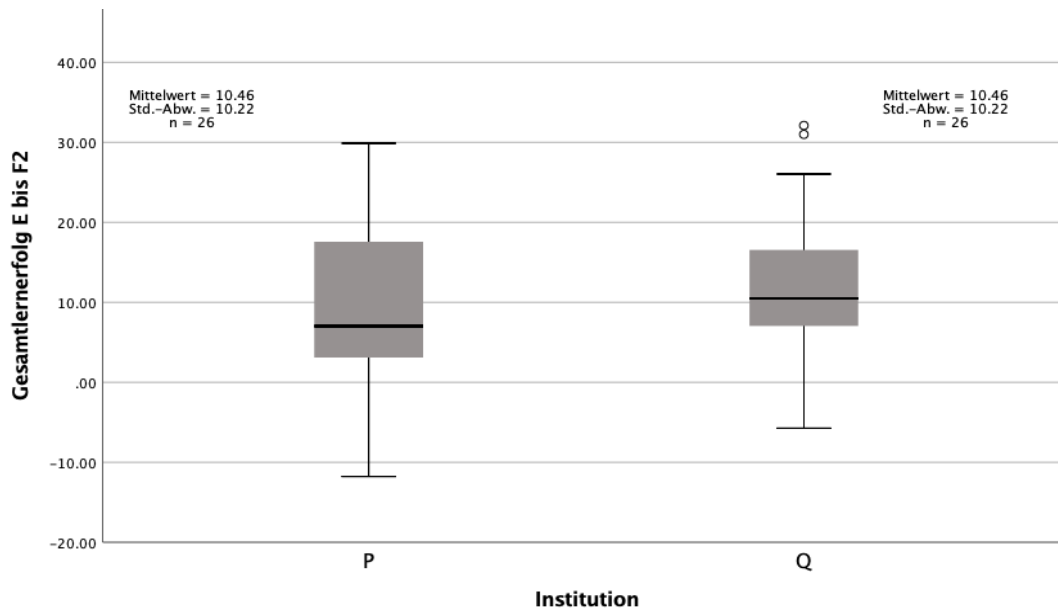


Abb. 27: Gesamtlernerfolg differenziert nach Institution

Bei genauerer Betrachtung der Kurslevel fällt auf, dass Institution Q in zwei von drei Kursniveaus einen höheren Lernzuwachs aufweist, wie aus Tab. 44 ersichtlich ist.

Die Durchführung einer ANOVA zeigt, dass das **Kurslevel** einen signifikanten Einfluss auf den Gesamtlernerfolg hat ($F(2,54) = 5.48$, $p = .007$, partielles $\eta^2 = .17$). 16.9 % (korr. R^2) der Streuung können durch das Kurslevel erklärt werden, mit einer starken Effektstärke ($\eta^2 = .17$). Post-hoc-Tests nach Bonferroni-Korrektur zeigen, dass sich das hohe Kurslevel signifikant vom tiefen Kurslevel ($p = .008$) unterscheidet. Auffällig ist, dass das tiefste Kurslevel den höchsten Mittelwert im Gesamtlernerfolg verzeichnet ($M = 15.06$, $SD = 10.67$).

Kurslevel tief, Institution Q	Mittelwert	17.40	Kurslevel tief, Institution P	Mittelwert	11.83
	Std.-Abweichung	11.69		Std.-Abweichung	8.78
	n	11		n	8
Kurslevel mittel, Institution Q	Mittelwert	9.30	Kurslevel mittel, Institution P	Mittelwert	17.30
	Std.-Abweichung	1.82		Std.-Abweichung	9.27
	n	9		n	8
Kurslevel hoch, Institution Q	Mittelwert	8.62	Kurslevel hoch, Institution P	Mittelwert	3.90
	Std.-Abweichung	6.18		Std.-Abweichung	8.49
	n	11		n	10

Tab. 44: Gesamtlernerfolg differenziert nach Kurslevel

Die Variable **Geschlecht** zeigt, dass weibliche Kursteilnehmende von der Einstufung bis zur zweiten Follow-up-Messung sich um rund 14 Punkte verbesserten, wohingegen männliche Lernende rund 10 Prozentpunkte zulegten, d. h. weibliche Teilnehmende weisen hier eine stärkere Lernprogression auf. Statistisch betrachtet ist der Gesamtlernerfolg jedoch nicht signifikant vom Geschlecht abhängig ($t(29,7) = -1.44$, $p = .161$), unter Berücksichtigung der Varianzheterogenität $F(1,55) = 11.339$, $p = .001$).

Die Variable **Alter** wird beim Gesamtlernerfolg anhand von Clustern analysiert: bis 19 Jahre, 20 bis 29 Jahre, 30 bis 39 Jahre, 40 bis 49 Jahre und über 50, wobei zwei Personen 51 und ein Teilnehmer 19 Jahre alt waren. Bei der Betrachtung dieser Variable fällt auf, dass der höchste Prozentpunktezuwachs mit fast 31 Prozentpunkten bei der 19-jährigen Person festgestellt werden kann, gefolgt von den zwei Teilnehmenden über 50 Jahren mit 16 Prozentpunkten ($SD = 9.41$, $n = 2$). Personen zwischen 20 und 29 Jahren haben den dritthöchsten Zuwachs im Umfang von etwas über 11 Prozentpunkten ($SD = 8.68$, $n = 28$), gefolgt von den 30- bis 39-Jährigen mit 11 Prozentpunkten ($SD = 9.48$, $n = 16$). Den tiefsten Lernzuwachs weisen Personen zwischen 40 und 49 Jahren mit 8 Prozentpunkten ($SD = 10.05$, $n = 10$) auf. Die Frage nach der statistischen Signifikanz wurde mittels ANOVA geprüft, die zeigt, dass die Unterschiede zwischen den Altersclustern statistisch nicht signifikant sind ($F(3,52) = .163$, $p = .181$, partielles $\eta^2 = .11$).

Den höchsten Lernzuwachs konnte der eine Teilnehmende mit der **Erstsprache** Saho mit fast 15 Prozentpunkten erreichen. Personen mit Tigrinya als L1 haben durchschnittlich den zweithöchsten Erfolg mit fast 13 Prozentpunkten erreicht ($SD = 9.57$, $n = 37$), gefolgt von den Personen mit Persisch als L1 mit knapp über 12 Prozentpunkten ($SD = 12.37$, $n = 2$). In der Mitte befinden sich Personen mit Erstsprache Arabisch mit knapp 12 Prozentpunkten ($SD = 6.52$, $n = 7$). Den zweittiefsten Wert erzielten Teilnehmende mit Kurdisch als Erstsprache mit fast 5 Prozentpunkten ($SD = 4.06$, $n = 5$). Der geringste Zuwachs ist bei Personen mit Tibetisch als Erstsprache mit durchschnittlich -4 Punkten ($SD = 10.51$, $n = 2$) zu verzeichnen. Die ANOVA zeigt jedoch, dass die Variable Erstsprache keinen signifikanten Einfluss auf den Gesamtlernerfolg hat ($F(5,50) = 1.78$, $p = .122$, partielles $\eta^2 = .18$).

Bei der Beherrschung des **Wortschatzes** in der Erstschriftsprache fällt auf, dass sich die Personen aus dem B- und C-Niveaubereich nur geringfügig unterscheiden. So zeigt die Gruppe im Bereich des B-Niveaus mit durchschnittlich 11 die meisten Prozentpunkte ($SD = 7.95$, $n = 11$), gefolgt von Lernenden im C-Niveaubereich mit fast 12 Prozentpunkten ($SD = 9.11$, $n = 37$). Teilnehmer mit den tiefsten Wortschatzkenntnissen im A2-Bereich in der Erstschriftsprache zeigten den geringsten Gesamtlernerfolg mit 9 Prozentpunkten ($SD = 12.49$, $n = 9$). Um die Mittelwerte zu vergleichen, wurde auch hier eine ANOVA durchgeführt, die keinen statistischen Zusammenhang erkennen lässt ($F(2,54) = .32$, $p = .727$, partielles $\eta^2 = .01$). Auch bezüglich der **formalen Richtigkeit** in der Erstschriftsprache sind nur geringe Unterschiede hinsichtlich der Niveaus erkennbar: Teilnehmer im A2-Bereich erzielten einen

Lernzuwachs von fast 10 Prozentpunkten ($SD = 10.34$, $n = 19$) und Personen mit einer Bewertung im B-Niveau von 11 Prozentpunkten ($SD = 8.78$, $n = 21$). Den höchsten Gesamtlernfortschritt mit 12 Prozentpunkten weisen Teilnehmende mit einer Bewertung im C-Niveaubereich auf ($SD = 9.38$, $n = 17$). Auch hier zeigt sich nach Durchführung einer ANOVA, dass die Varianz in der unabhängigen Variable hinsichtlich der formalen Richtigkeit und dem Gesamtlernerfolg nicht statistisch signifikant ist ($F(2,54) = .30$, $p = .741$, partielles $\eta^2 = .01$).

Auffällig ist, dass Teilnehmer ohne **Fremdsprachenkenntnisse** mit 13 Prozentpunkten ($SD = 9.28$, $n = 44$) einen deutlich höheren Lernzuwachs zeigen als mehrsprachige Teilnehmende mit fast 5 Prozentpunkten ($SD = 6.81$, $n = 13$, s. Abb. 28). Ein t-Test ergibt, dass dieser Unterschied statistisch signifikant ist ($t(55) = 2.94$, $p = .005$) und zwar mit einem starken Effekt ($d = .95$). Werden diese Unterschiede etwas genauer betrachtet, dann schnitten diejenigen, die eine Sprache mit einem Roman-alphabet beherrschen ($M = 6.97$, $SD = 4.12$, $n = 5$) besser ab, als Personen, die einer Sprache mit einem non-Roman-alphabet mächtig sind ($M = 5.89$, $SD = 5.52$, $n = 7$). Ein Teilnehmer gab an, Kenntnisse in Chinesisch (non-alphabet) zu haben; dieser erzielte den tiefsten Lernzuwachs von -11.77 Prozentpunkten, wobei dieser Wert als Ausreisser zu klassifizieren ist. Die einfaktorielle Varianzanalyse bestätigt, dass Fremdsprachenkenntnisse einen signifikanten Einfluss auf den Gesamtlernfortschritt haben ($F(1,55) = 8.67$, $p = .005$, partielles $\eta^2 = .14$); ein Post-hoc-Bonferroni-Test kann jedoch aufgrund der zu geringen Anzahl an Fällen nicht durchgeführt werden.

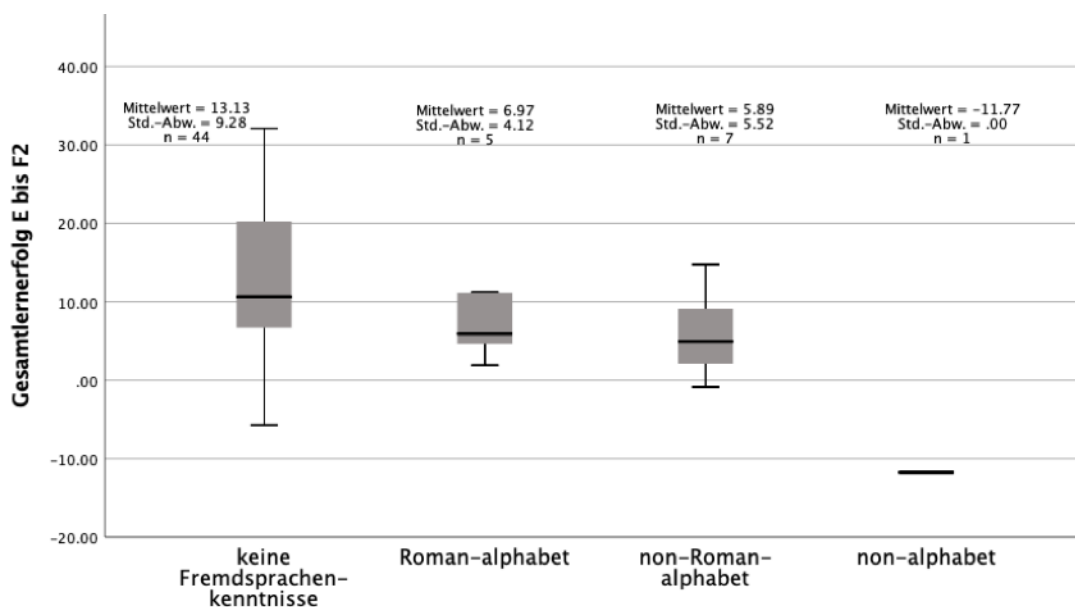


Abb. 28: Gesamtlernerfolg differenziert nach Fremdsprachenkenntnissen

Zur Analyse des Effekts der **Anzahl Schuljahre** auf den Gesamtlernerfolg wird auf die bereits gebildeten Cluster zurückgegriffen: 1 bis 3 Schuljahre ($n = 5$), 4 bis 6 Schuljahre ($n = 13$), 7 bis 9 Schuljahre ($n = 29$) und 10 bis 12 Schuljahre ($n = 9$). Den durchschnittlich höchsten Prozentpunktezuwachs von 17 weist der Lernende ohne Schulerfahrung auf, gefolgt von jeweils rund 12 Punkten ($SD = 8.83$) bei den Teilnehmenden mit 7 bis 9 Schuljahren und 10 bis 12 Schuljahren ($SD = 10.90$). Den vierthöchsten respektive zweitniedrigsten Wert von durchschnittlich 10 Punkten erzielten Personen mit 1 bis 3 Schuljahren ($SD = 16.04$) und Teilnehmende mit 4 bis 6 Schuljahren erreichten den geringsten Prozentpunktwert von 8 ($SD = 7.02$). Die ANOVA zeigt jedoch keinen signifikanten Einfluss der Anzahl Schuljahre auf den Gesamtlernerfolg ($F(3,52) = .47, p = .755$, partielles $\eta^2 = .04$).

Was die Variable **Arbeit** im Heimatland betrifft, fällt auf, dass Teilnehmer ohne Arbeitserfahrung fast 14 und damit mehr Prozentpunkte im Gesamtlernerfolg ($SD = 10.37, n = 22$) erzielten als Lernende mit Arbeitserfahrung (9 Prozentpunkte, $SD = 8.41, n = 35$).

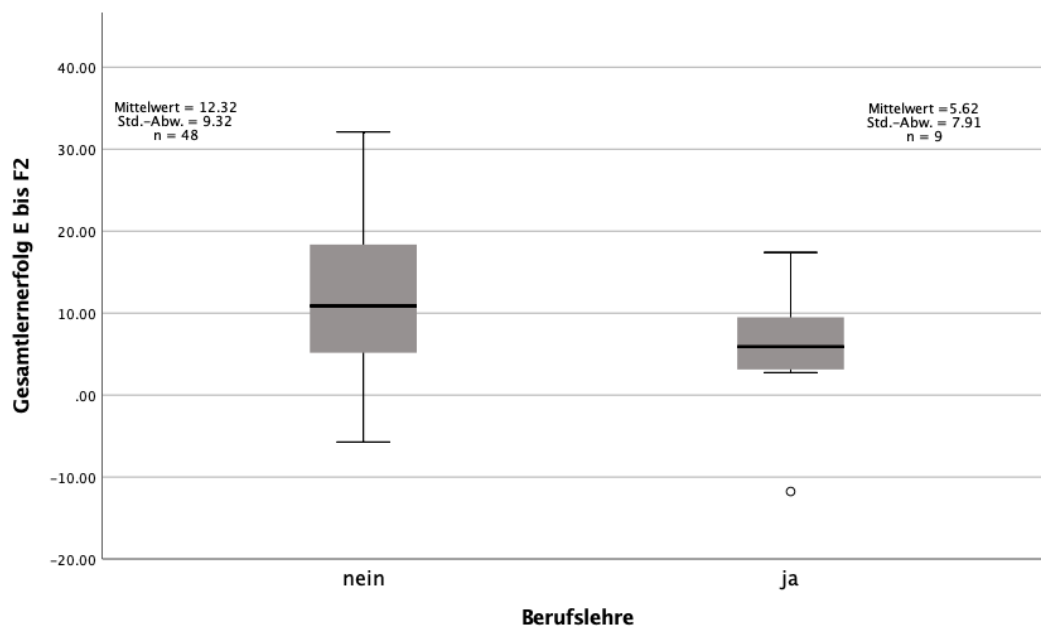


Abb. 29: Gesamtlernerfolg differenziert nach Berufslehre

Ähnliches zeigt sich hinsichtlich **Berufslehre** (s. Abb. 29): Teilnehmer mit einer Berufslehre oder einem äquivalenten Abschluss aus dem Herkunftsland schnitten mit einem durchschnittlichen Lernzuwachs von fast sechs Prozentpunkten ($SD = 7.91, n = 9$) deutlich schlechter ab als Teilnehmer ohne Berufslehre mit 12 Prozentpunkten ($SD = 9.32, n = 48$). Ein Teilnehmer gab an, in seinem Heimatland die **Universität** besucht zu haben. Dieser erzielte einen höheren Lernzuwachs als die restlichen Teilnehmenden ($M = 21.72$ im Gegensatz zu

$M = 11.07, SD = 9.36$). t-Tests zeigen lediglich für die Variable Berufslehre ($t(55) = 2.02, p = .048$) einen signifikanten Unterschied, mit einer mittleren Effektstärke ($d = .75$).

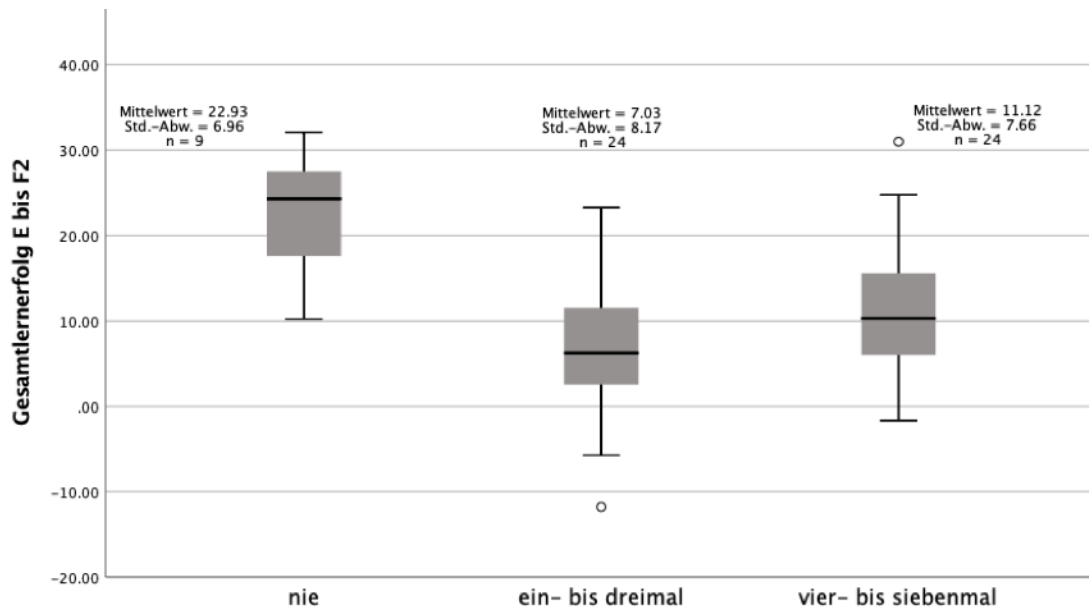


Abb. 30: Gesamtlernerfolg differenziert nach Deutschgebrauch

Bei Betrachtung der Lernprogression vom Einstufungstest bis hin zur zweiten Follow-up-Messung zeigt sich (s. Abb. 30), dass Teilnehmende, die ausserhalb des Unterrichts kein **Deutsch benutzten**, den höchsten Lernzuwachs aufwiesen ($M = 22.93, SD = 6.96, n = 9$). Bei Lernenden, die vier- bis siebenmal pro Woche auf Deutsch kommunizierten, kann der zweithöchste Zuwachs beobachtet werden, ($M = 11.12, SD = 7.66, n = 24$), gefolgt von Personen, die ein bis dreimal pro Woche Deutsch benutzten ($M = 7.03, SD = 8.17, n = 24$). Diese Befunde sind statistisch hochsignifikant ($F(2,54) = 13.66, p = .000$, partielles $\eta^2 = .34$) mit einer mittleren Effektstärke ($d = .52$).

Die hier untersuchten Teilnehmer, für jene es ihr erster **Deutschkurs** war, haben mit fast 19 Prozentpunkten ($SD = 9.99, n = 11$) zwischen Einstufungstest und der zweiten Follow-up-Messung deutlich mehr zugelegt als Teilnehmer, die vorgängig bereits einen oder mehrere Deutschkurs(e) besucht haben mit 9 Prozentpunkten ($SD = 8.41, n = 46$, s. Abb. 31). Ein t-Test zeigt, dass Lernende ohne vorgängig besuchten Deutschkurs ($M = 18.65, SD = 9.99$) einen höheren Gesamtlernerfolg aufweisen als Teilnehmer, die bereits einen Deutschkurs absolviert haben ($M = 9.49, SD = 8.41$), wobei sich der Mittelwertunterschied als signifikant erweist ($t(55) = 3.13, p = .003$). Die Effektstärke nach Cohen liegt bei $d = 1.07$ und entspricht einem starken Effekt.

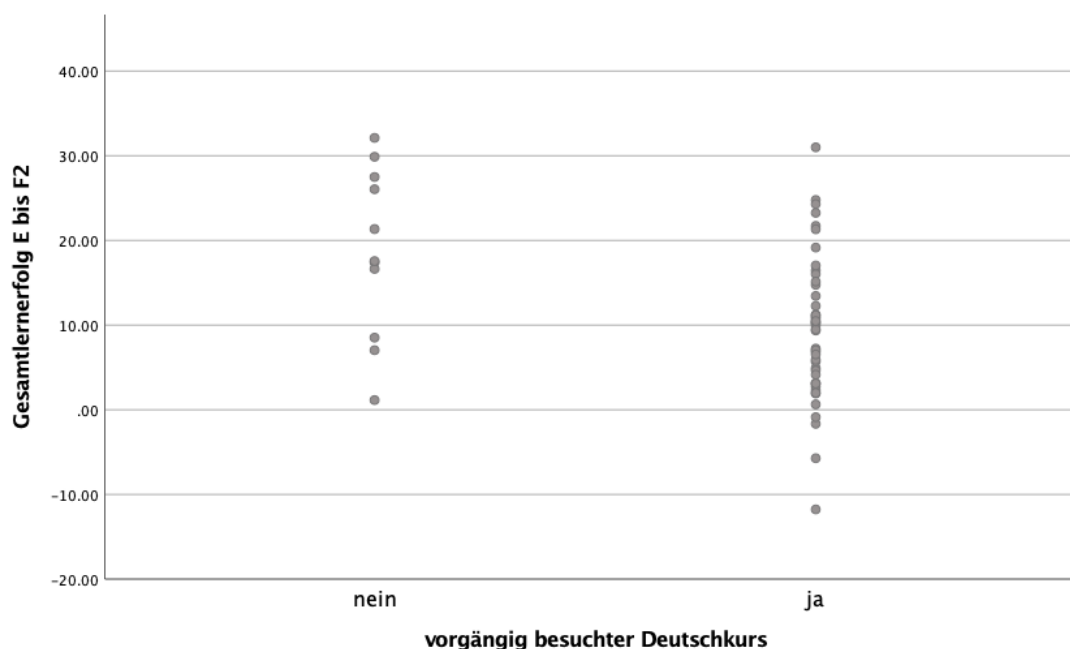


Abb. 31: Gesamtlernerfolg differenziert nach vorgängig besuchttem Deutschkurs

Teilnehmende, die mit der **Silbenmethode** unterrichtet wurden, verzeichneten einen tieferen Punktezuwachs (10 Prozentpunkte, $SD = 10.20$, $n = 30$) als Lernende ohne zusätzlichem Fokus auf dieser Methode mit 12 Prozentpunkten ($SD = 8.49$, $n = 27$). Statistisch betrachtet ist dieser Unterschied aber nicht bedeutsam ($t(55) = .67$, $p = .507$).

6.3.2 Die Treatments im Kursverlauf

An dieser Stelle soll der Lernerfolg der Teilnehmenden mit/ohne Silbenmethode im Kursverlauf betrachtet werden. Hierzu wird der Lernerfolg in den einzelnen Kursphasen herangezogen, um zu sehen, ob ein kontinuierlicher Lernanstieg durch die Testpersonen erreicht wurde und ob dieser je nach Treatment oder Kurslevel unterschiedlich ausfällt (s. Tab. 45). Der Lernerfolg der einzelnen Kursphasen wird analog zum Gesamtlernerfolg (s. Abschn. 5.3.1) berechnet, d. h. vom durchschnittlichen Resultat aller Teilnehmenden am Ende der untersuchten Kursphase wird das durchschnittliche Resultat am Anfang derselben Kursphase abgezogen. Das bedeutet demnach, dass sich beispielsweise der Lernerfolg in der ersten Kursphase definiert als [Punktzahl Follow-up 1] minus [Punktzahl Einstufung].

	Einstufung	Follow-up 1	Follow-up 2
Mittelwert	41.54	46.97	52.80
Std.-Abweichung	16.72	12.21	12.86
N	57	57	57

Tab. 45: Mittelwerte des Lernerfolgs zu allen drei Messzeitpunkten

Eine Varianzanalyse mit Messwiederholung – unter entsprechender Greenhouse-Geisser-Korrektur aufgrund des signifikant ausgefallenen Mauchly-Tests auf Sphärizität (Mauchly- $W(2)$

= .576, $p = .000$) – zeigt, dass sich die pro Messzeitpunkt total erreichten Punktwerte hochsignifikant unterscheiden ($F(1.4, 78.64) = 18.934, p = .000, \eta^2 = .25$).

	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 1 (E zu F1)	Kurslevel tief ohne Silbenmethode	8	5.70	5.59	1.98
	Kurslevel tief mit Silbenmethode	11	11.70	7.70	2.32
	Kurslevel mittel ohne Silbenmethode	8	11.57	6.41	2.27
	Kurslevel mittel mit Silbenmethode	9	6.97	4.00	1.33
	Kurslevel hoch ohne Silbenmethode	11	.28	6.56	1.98
	Kurslevel hoch mit Silbenmethode	10	-2.30	6.91	2.18

Tab. 46: Lernerfolg erste Phase (Einstufung bis Follow-up 1)

Der Lernerfolg der **ersten Phase** (s. Tab. 46), d. h. von der Einstufung bis zur ersten Follow-up-Messung, zeigt, dass die Silbenmethode in der Anfangsphase, d. h. mit der Einführung in das lateinische Schriftsystem, einen höheren Mittelwert im tiefen Kurslevel aufweist. In den beiden höheren Kursen ergibt sich für die Silbenmethode jedoch ein geringerer Wert als für die Kurse ohne Silbenmethode. Nach Durchführung eines t-Tests zeigt sich, dass die Varianzen statistisch nicht signifikant ($t(17) = -1.87, p = .079$) sind.

	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 2 (F1 zu F2)	Kurslevel tief ohne Silbenmethode	8	6.13	7.73	2.73
	Kurslevel tief mit Silbenmethode	11	5.70	6.16	1.86
	Kurslevel mittel ohne Silbenmethode	8	5.73	5.32	1.88
	Kurslevel mittel mit Silbenmethode	9	2.32	3.80	1.27
	Kurslevel hoch ohne Silbenmethode	11	8.34	2.76	.83
	Kurslevel hoch mit Silbenmethode	10	6.20	3.58	1.13

Tab. 47: Lernerfolg zweite Phase (Follow-up 1 bis Follow-up 2)

In der **zweiten Phase** (s. Tab. 47) d. h. von der ersten Follow-up-Messung bis zur zweiten, wird deutlich, dass die Silbenmethode durchgängig einen geringeren Mittelwert erzielt. Ein t-Test macht deutlich, dass sich die Mittelwerte in keinem Kurslevel auf einem Signifikanzniveau von .05 unterscheiden ($t(17) = -.13, p = .895$).

	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Gesamtlernerfolg (E zu F2)	Kurslevel tief ohne Silbenmethode	8	11.83	8.78	3.11
	Kurslevel tief mit Silbenmethode	11	17.40	11.69	3.52
	Kurslevel mittel ohne Silbenmethode	8	17.30	9.27	3.28
	Kurslevel mittel mit Silbenmethode	9	9.30	1.82	.61
	Kurslevel hoch ohne Silbenmethode	11	8.62	6.18	1.86
	Kurslevel hoch mit Silbenmethode	10	3.90	8.49	2.69

Tab. 48: Gesamtlernerfolg (Einstufung bis Follow-up 2)

Hinsichtlich des **Gesamtlernerfolgs** (s. Tab. 48), d. h. von der Einstufung bis zur zweiten Follow-up-Messung, verzeichnen die Kurse mit der Silbenmethode lediglich im tiefen Kurslevel einen höheren Gesamtlernerfolg als die Kurse ohne Silbenmethode. Signifikant sind

die Unterschiede aber nur im mittleren Kurslevel, zugunsten der Gruppe ohne Silbenmethode ($t(15) = 2.544$, $p = .022$), unter Berücksichtigung der Varianzheterogenität $F(1,15) = 7.866$, $p = .013$).

6.3.3 Die Treatments bei den forschergenerierten Tests

Beim **Elicited Imitation** zeigt sich, dass Teilnehmende aus dem tiefen Kurslevel mit Silbenmethode durchgehend einen höheren Lernzuwachs erreichen konnten. Im mittleren Kurslevel weist der Kurs mit Silbenmethode beim Lernerfolg 1 (s. Tab. 49) und beim Gesamtlernerfolg höhere (s. Tab. 51), beim Lernerfolg 2 hingegen tiefere Werte auf (s. Tab. 50). Einzig im hohen Kurslevel erreichte der Kurs mit der Silbenmethode durchgängig deutlich tiefere Werte als sein Pendant. t-Tests offenbaren, dass lediglich im hohen Kurslevel der Gesamtlernerfolg, d. h. von der Einstufung bis zum Follow-up 2, statistisch signifikant ist ($p = .024$).

Elicited Imitation	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 1 (E zu F1)	Kurslevel tief ohne Silbenmethode	8	5.42	11.90	4.21
	Kurslevel tief mit Silbenmethode	11	10.08	10.28	3.10
	Kurslevel mittel ohne Silbenmethode	8	.83	5.23	1.85
	Kurslevel mittel mit Silbenmethode	9	4.91	5.29	1.76
	Kurslevel hoch ohne Silbenmethode	11	1.97	9.33	2.81
	Kurslevel hoch mit Silbenmethode	10	-1.50	4.19	1.33

Tab. 49: Lernerfolg erste Phase (Einstufung bis Follow-up 1), Elicited Imitation

Elicited Imitation	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 2 (F1 zu F2)	Kurslevel tief ohne Silbenmethode	8	1.67	3.15	1.11
	Kurslevel tief mit Silbenmethode	11	3.18	3.35	1.01
	Kurslevel mittel ohne Silbenmethode	8	1.56	7.09	2.51
	Kurslevel mittel mit Silbenmethode	9	.19	3.14	1.05
	Kurslevel hoch ohne Silbenmethode	11	.08	6.05	1.82
	Kurslevel hoch mit Silbenmethode	10	-2.67	5.06	1.60

Tab. 50: Lernerfolg zweite Phase (Follow-up 1 bis Follow-up 2), Elicited Imitation

Elicited Imitation	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Gesamtlernerfolg (E zu F2)	Kurslevel tief ohne Silbenmethode	8	7.08	11.06	3.91
	Kurslevel tief mit Silbenmethode	11	13.26	9.06	2.73
	Kurslevel mittel ohne Silbenmethode	8	2.40	10.47	3.70
	Kurslevel mittel mit Silbenmethode	9	5.09	7.22	2.41
	Kurslevel hoch ohne Silbenmethode	11	2.05	6.55	1.98
	Kurslevel hoch mit Silbenmethode	10	-4.17	4.76	1.51

Tab. 51: Gesamtlernerfolg (Einstufung bis Follow-up 2), Elicited Imitation

Im **Hörverstehen** weist der Kurs auf tiefem Niveau mit Silbenmethode von der Einstufung bis zum Follow-up 1 (s. Tab. 52) und von der Einstufung bis zum Follow-up 2 (s. Tab. 54) einen höheren durchschnittlichen Mittelwert auf als der Kurs ohne zusätzlichem Fokus auf der Silbenmethode. Im hohen Kurslevel wird mit der Silbenmethode in der zweiten Kursphase (s.

Tab. 53) ein geringfügig höherer Mittelwert erzielt, das mittlere Kurslevel weist für die Bedingung ‚mit Silbenmethode‘ hingegen beim Lernerfolg 1 und beim Gesamtlernerfolg einen geringeren Mittelwert auf. t-Tests können diese Werte belegen: Die Varianzen im tiefen Kurslevel beim ersten Lernerfolg ($p = .045$) und beim Gesamtlernerfolg ($p = .033$) unterscheiden sich auf einem .05-Signifikanzniveau. Dasselbe gilt für das mittlere Kurslevel ($p = .019$) und ($p = .022$).

Hörverstehen	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 1 (E zu F1)	Kurslevel tief ohne Silbenmethode	8	.29	21.37	7.56
	Kurslevel tief mit Silbenmethode	11	23.02	23.46	7.07
	Kurslevel mittel ohne Silbenmethode	8	26.68	19.89	7.03
	Kurslevel mittel mit Silbenmethode	9	4.32	14.99	5.00
	Kurslevel hoch ohne Silbenmethode	11	-8.88	12.90	3.89
	Kurslevel hoch mit Silbenmethode	10	-12.54	12.94	4.09

Tab. 52: Lernerfolg erste Phase (Einstufung bis Follow-up 1), Hörverstehen

Hörverstehen	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 2 (F1 zu F2)	Kurslevel tief ohne Silbenmethode	8	12.50	24.68	8.72
	Kurslevel tief mit Silbenmethode	11	11.97	21.20	6.39
	Kurslevel mittel ohne Silbenmethode	8	.00	13.78	4.87
	Kurslevel mittel mit Silbenmethode	9	.00	16.01	5.34
	Kurslevel hoch ohne Silbenmethode	11	11.48	10.74	3.24
	Kurslevel hoch mit Silbenmethode	10	11.58	10.47	3.31

Tab. 53: Lernerfolg zweite Phase (Follow-up 1 bis Follow-up 2), Hörverstehen

Hörverstehen	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Gesamtlernerfolg (E zu F2)	Kurslevel tief ohne Silbenmethode	8	12.79	12.92	4.57
	Kurslevel tief mit Silbenmethode	11	34.98	24.57	7.41
	Kurslevel mittel ohne Silbenmethode	8	26.68	19.00	6.72
	Kurslevel mittel mit Silbenmethode	9	4.32	17.15	5.72
	Kurslevel hoch ohne Silbenmethode	11	2.61	13.11	3.95
	Kurslevel hoch mit Silbenmethode	10	-.96	10.84	3.43

Tab. 54: Gesamtlernerfolg (Einstufung bis Follow-up 2), Hörverstehen

Bei der **Oral Reading Fluency** weist lediglich das tiefe Kurslevel mit Silbenmethode in der zweiten Kursphase (s. Tab. 56) einen höheren Mittelwert auf als der Kurs ohne Silbenmethode. Ansonsten können stets die Kurse ohne zusätzlichem Fokus auf der Silbenmethode einen höheren Mittelwert verzeichnen (s. Tab. 55 und Tab. 57). Nach Durchführung von t-Tests können jedoch keine statistisch signifikanten Unterschiede festgestellt werden ($p \geq .05$).

Oral Reading Fluency (ORF)	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 1 (E zu F1)	Kurslevel tief ohne Silbenmethode	8	11.75	8.21	2.90
	Kurslevel tief mit Silbenmethode	11	2.64	19.30	5.82
	Kurslevel mittel ohne Silbenmethode	8	9.13	13.02	4.60
	Kurslevel mittel mit Silbenmethode	9	8.56	8.03	2.68
	Kurslevel hoch ohne Silbenmethode	11	1.91	11.57	3.49
	Kurslevel hoch mit Silbenmethode	10	-.70	11.94	3.78

Tab. 55: Lernerfolg erste Phase (Einstufung bis Follow-up 1), ORF

Oral Reading Fluency (ORF)	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 2 (F1 zu F2)	Kurslevel tief ohne Silbenmethode	8	4.25	5.71	2.02
	Kurslevel tief mit Silbenmethode	11	6.18	4.75	1.43
	Kurslevel mittel ohne Silbenmethode	8	7.38	7.67	2.72
	Kurslevel mittel mit Silbenmethode	9	2.22	4.82	1.61
	Kurslevel hoch ohne Silbenmethode	11	10.73	7.82	2.36
	Kurslevel hoch mit Silbenmethode	10	4.40	6.02	1.90

Tab. 56: Lernerfolg zweite Phase (Follow-up 1 bis Follow-up 2), ORF

Oral Reading Fluency (ORF)	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Gesamtlernerfolg (E zu F2)	Kurslevel tief ohne Silbenmethode	8	16.00	11.46	4.05
	Kurslevel tief mit Silbenmethode	11	8.82	21.72	6.55
	Kurslevel mittel ohne Silbenmethode	8	16.50	18.28	6.46
	Kurslevel mittel mit Silbenmethode	9	10.78	6.51	2.17
	Kurslevel hoch ohne Silbenmethode	11	12.64	11.77	3.55
	Kurslevel hoch mit Silbenmethode	10	3.70	14.73	4.66

Tab. 57: Gesamtlernerfolg (Einstufung bis Follow-up 2), ORF

Beim **Diktat** zeigt sich, dass die Kurse mit zusätzlichem Fokus auf der Silbenmethode im tiefen und mittleren Kurslevel zwischen Einstufung und Follow-up 1 (s. Tab. 58) höhere durchschnittliche Mittelwerte aufweisen. In der Kursphase zwischen Follow-up 1 und Follow-up 2 (s. Tab. 59) weist lediglich das hohe Kurslevel mit Silbenmethode einen geringfügig höheren durchschnittlichen Wert auf. Bei der dritten Lernphase, d. h. von der Einstufung bis zum Follow-up 2 (s. Tab. 60) ergibt sich nur im tiefen Kurslevel mit Silbenmethode ein geringfügig höherer Mittelwert. Die Unterschiede sind statistisch nicht signifikant ($p \geq .05$).

Diktat	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 1 (E zu F1)	Kurslevel tief ohne Silbenmethode	8	5.33	7.53	2.66
	Kurslevel tief mit Silbenmethode	11	11.08	7.25	2.19
	Kurslevel mittel ohne Silbenmethode	8	9.63	7.88	2.79
	Kurslevel mittel mit Silbenmethode	9	10.11	9.33	3.11
	Kurslevel hoch ohne Silbenmethode	11	6.12	7.63	2.30
	Kurslevel hoch mit Silbenmethode	10	5.55	13.15	4.16

Tab. 58: Lernerfolg erste Phase (Einstufung bis Follow-up 1), Diktat

Diktat	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Lernerfolg 2 (F1 zu F2)	Kurslevel tief ohne Silbenmethode	8	6.10	10.71	3.79
	Kurslevel tief mit Silbenmethode	11	1.48	5.77	1.74
	Kurslevel mittel ohne Silbenmethode	8	13.98	8.55	3.02
	Kurslevel mittel mit Silbenmethode	9	6.89	6.45	2.15
	Kurslevel hoch ohne Silbenmethode	11	11.08	10.10	3.05
	Kurslevel hoch mit Silbenmethode	10	11.48	15.62	4.94

Tab. 59: Lernerfolg zweite Phase (Follow-up 1 bis Follow-up 2), Diktat

Diktat	Kurslevel	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
Gesamtlernerfolg (E zu F2)	Kurslevel tief ohne Silbenmethode	8	11.44	13.67	4.83
	Kurslevel tief mit Silbenmethode	11	12.56	9.78	2.95
	Kurslevel mittel ohne Silbenmethode	8	23.60	7.98	2.82
	Kurslevel mittel mit Silbenmethode	9	17.00	8.61	2.87
	Kurslevel hoch ohne Silbenmethode	11	17.18	8.28	2.50
	Kurslevel hoch mit Silbenmethode	10	17.03	17.06	5.39

Tab. 60: Gesamtlernerfolg (Einstufung bis Follow-up 2), Diktat

Zusammenfassend lässt sich bei der Elicited Imitation für das hohe Kurslevel zwischen Einstufung und Follow-up 2 ein geringerer und gar negativer Gesamtlernerfolg für den Unterricht mit Silbenmethode nachweisen. Beim Hörverstehen wirkt sich die Silbenmethode im mittleren Kurslevel zwischen Einstufung und Follow-up 1 sowie zwischen Einstufung und Follow-up 2 ebenfalls negativ auf den Lernerfolg aus. Im tiefen Kursniveau führt der Unterricht mit der Silbenmethode im Hörverstehen hingegen zu einer erhöhten Lernprogression. Alle anderen Zusammenhänge sind statistisch nicht signifikant.

6.3.4 Regression

Das unten stehende Regressionsmodell³⁵ wurde, analog zur Untersuchung der telc-Resultate, theoriegeleitet erstellt. Der Gesamtlernfortschritt in Prozentpunkten, d. h. von der Einstufung bis zur zweiten Follow-up-Messung nach 200 Kurslektionen, wurde als abhängige Variable gewählt. Er wird als Differenz zwischen der Prozentzahl korrekter Antworten beim Einstufungstest und dem Resultat beim zweiten Follow-up-Test operationalisiert (s. Abschn. 5.3.1). Die bereits eingeführten strukturellen Merkmale dienen auch hier als Regressoren. Um Probleme der Multikollinearität zu vermeiden, wurden hier ebenfalls die Variablen *Kurslevel tief*, das Alterscluster *20 bis 29*, die *L1 Tigrinya, 7 bis 9 Schuljahre* und *kein Deutschgebrauch* ausserhalb des Unterrichts als Referenzkategorien gewählt.

Das verwendete Regressionsmodell weist ein korr. R^2 von .483 ($R^2 = .787$) auf, was darauf hinweist, dass 48 % der Varianz des Gesamtlernfortschrittes durch die untersuchten Variablen erklärt werden können. Wie aus dem Regressionsmodell hervorgeht (s. Tab. 61), verfügt unter Konstanthaltung der übrigen Faktoren lediglich eine Variable über einen signifikanten Koeffizienten. Es handelt sich um die Variable **Alter** ≤ 19. Die 19-jährige Probandin erzielte einen massiv höheren Lernzuwachs als die anderen Kursteilnehmenden ($B = 24.33$, $p = .012$).

³⁵ Die Voraussetzungen der linearen Regressionsanalyse wurden auch hier überprüft und konnten eingehalten werden: Intervallskala der UV und AV, Linearität des Zusammenhangs, Unabhängigkeit und Normalverteilung des Fehlerwertes sowie die Gauss-Markov-Annahmen 1–5.

Da es sich hier aber um einen Einzelfall handelt, ist dieses Resultat nicht generalisierbar. Die Variable 4–6 *Schuljahre* liegt knapp ausserhalb des statistisch signifikanten Schwellenwertes. Aus der linearen Regression lassen sich somit, trotz Signifikanz des Gesamtmodells ($p = .009$), wenig Informationen bezüglich für den Gesamtlernerfolg relevante Faktoren gewinnen.

Modellzusammenfassung						
R	R-Quadrat	Korrigiertes R-Quadrat		Standardfehler des Schätzers		
.887a	.787	.483		6.74732		
Koeffizienten ^a						
	Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Sig.	
	Regressions- koeffizientenB	Std.-Fehler	Beta			
(Konstante)	27.45	9.59		2.86	.009	
Institution	−2.38	2.81	−.13	−.85	.406	
Kurslevel mittel	−3.09	3.12	−.15	−.99	.333	
Kurslevel hoch	−5.32	3.70	−.28	−1.44	.163	
Silbenmethode	−3.32	2.63	−.18	−1.26	.220	
Geschlecht	−.39	3.49	−.02	−.11	.912	
Alter ≤ 19	24.33	8.96	.34	2.72	.012*	
Alter 30 bis 39	−3.57	3.24	−.17	−1.10	.281	
Alter 40 bis 49	−2.57	3.91	−.11	−.66	.517	
Alter ≥ 50	−.81	6.91	−.02	−.12	.908	
L1 Saho	12.11	10.50	.17	1.15	.261	
L1 Arabisch	7.46	4.85	.26	1.54	.138	
L1 Kurdisch	−3.77	4.84	−.12	−.78	.444	
L1 Persisch	−13.31	8.73	−.32	−1.53	.141	
L1 Tamil	−9.70	8.22	−.19	−1.18	.250	
L1 Tibetisch	−10.07	8.05	−.20	−1.25	.224	
Wortschatz Erstschrift B1/B2	−6.21	6.48	−.26	−.96	.348	
Wortschatz Erstschrift C1/C2	−2.82	5.56	−.15	−.51	.617	
Formale Richtigkeit Erstschrift B1/B2	8.79	5.02	.46	1.75	.094	
Formale Richtigkeit Erstschrift C1/C2	5.39	4.96	.27	1.09	.288	
Fremdsprache	−6.61	3.39	−.29	−1.95	.063	
Schuljahre 0	13.40	10.59	.19	1.27	.219	
Schuljahre 1 bis 3	.10	5.17	.03	.19	.849	
Schuljahre 4 bis 6	−6.71	3.28	−.30	−2.05	.052	
Schuljahre 10 bis 12	.05	4.63	.00	.01	.991	
Berufslehre	−2.95	4.01	−.12	−.74	.470	
Universität	14.70	12.72	.21	1.16	.260	
Arbeit Heimatland	1.13	2.67	.06	.42	.677	
Kinder	.11	3.57	.01	.03	.976	
Gesundheitszustand	−.13	1.43	−.01	−.09	.928	
1–3x pro Woche Deutschgebrauch	−6.46	5.48	−.34	−1.18	.250	
4–7x pro Woche Deutschgebrauch	−2.49	4.92	−.13	−.51	.618	
vorgängig besuchter Deutschkurs	−5.50	4.01	−.23	−1.37	.184	
Kursabsenz	.050	.14	.06	.36	.722	

Tab. 61: Regressionsmodell Gesamtlernerfolg (Einstufung bis Follow-up 2)

* = .05-Signifikanzniveau, ** = .01-Signifikanzniveau

6.3.5 Überprüfung der Hypothesen

Auf Basis der vorangehenden Analysen zum Lernerfolg lassen sich nun die Hypothesen überprüfen.

H3. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im Lernerfolg.

- H3.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.
- H3.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.
- H3.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen.
- H3.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.

Die Prüfung der Hypothese H3 kann sich auf die bivariate Analyse in Abschnitt 6.3.1 und der Regressionsanalyse in 6.3.4 stützen. Auf die Unterhypothesen H3.1–H3.4 wird im Rahmen der Ausführungen in Abschnitt 6.3 und 6.3.3 näher eingegangen.

Obwohl in der bivariaten Analyse ein Unterschied zwischen den Mittelwerten der Treatmentgruppen beobachtet werden konnte, weist dieser keine statistische Signifikanz auf. Diese fehlende Signifikanz zeigt sich auch in der Regressionsanalyse, weshalb Hypothese H3 als falsifiziert gilt.

Was die Unterhypothesen H3.1–H3.4 betrifft, so sind bezüglich den Fertigkeiten Lesen und Schreiben keine Zusammenhänge erkennbar. Im **Sprechen** weist das Treatment mit der Silbenmethode für das hohe Kurslevel einen signifikant tieferen Gesamtlernerfolg auf. In der Fertigkeit **Hören** weist das tiefe Kurslevel in der ersten Kursphase (Einstufung bis Follow-up 1) und im Gesamtlernerfolg (Einstufung bis Follow-up 2) statistisch signifikant bessere Ergebnisse auf und das mittlere Kurslevel zeigt durchgehend signifikant schwächere Resultate. Auch wenn das zusätzliche Treatment weitere positive Einflüsse ergeben hat, weisen diese keine statistische Signifikanz auf und können demnach für die Beantwortung der Hypothesen nicht herangezogen werden. Die Hypothesen H3.1 und H3.2 müssen somit falsifiziert werden. Die Hypothesen H3.3 und H3.4 müssen in der formulierten Pauschalität ebenfalls zurückgewiesen werden.

H4. Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat und den Gesamtlernerfolg.

- H4.2 Die Anzahl der Schuljahre hat einen Einfluss auf den Gesamtlernerfolg.

In der bivariaten Analyse (s. Abschn. 6.3.1) weisen die Schuljahre-Cluster unterschiedliche Mittelwerte auf, die Unterschiede sind aber statistisch nicht signifikant. In der

Regressionsanalyse (s. Abschn. 6.3.4) sind die entsprechenden Variablen allesamt nicht signifikant. Hypothese H4.2 wird durch die vorliegenden Daten somit falsifiziert.

H5. Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das Resultat im telc-Test und den Gesamtlernerfolg.

H5.2 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst den Gesamtlernerfolg.

Zwar zeigen sich in der bivariaten Analyse bezüglich der Variablen L1 Wortschatz und L1 formaler Richtigkeit Mittelwertunterschiede, jedoch erwiesen sich diese statistisch als nicht signifikant (s. Abschn. 6.3.1). Auch anhand der Regressionsanalyse ist kein statistischer Zusammenhang zu erkennen (s. Abschn. 6.3.4). Hypothese H5.2 kann demnach falsifiziert werden.

6.4 Analyse der Kunstwörter

Zur Einstufung und bei beiden Follow-up-Messungen wurden jeweils die gleichen 30 Kunstwörter abgefragt, um die Gesamtkapazität der phonologischen Schleife, d. h. die Kapazität des Arbeitsgedächtnisses, der Teilnehmenden zu überprüfen (s. Abschn. 4.2.2.1). Die Wichtigkeit der phonologischen Schleife konnte u. a. beim Lesenlernen und im Wortschatzerwerb bei Kindern mit Deutsch als Mutter- oder Zweitsprache bestätigt werden (vgl. Ulrich 2016, S. 23). Aus diesem Grund soll hier untersucht werden, ob sich ein Zusammenhang zwischen der Silbenmethode und der Gesamtkapazität der phonologischen Schleife zeigt und inwiefern dieses für das Erreichen des A1-Niveaus von Belang ist. Für die Kapazität des Arbeitsgedächtnisses dient als Proxy der Punktezuwachs bei den Kunstwörtern von der Einstufung bis zum Follow-up 2. Das heisst, das bessere Abschneiden bei den Kunstwörtern beim Follow-up 2 wird mit dem Arbeitsgedächtnis in Verbindung gebracht.

	Silbenmethode	n	Mittelwert	Std.-Abw.	Standardfehler des Mittelwertes
E zu F2 alle Kurslevel	ohne Silbenmethode	27	3.19	7.40	1.42
	mit Silbenmethode	30	6.90	9.41	1.72
E zu F2 Kurslevel tief	ohne Silbenmethode	8	3.88	5.66	2.00
	mit Silbenmethode	11	4.70	11.26	3.39
E zu F2 Kurslevel mittel	ohne Silbenmethode	8	6.03	6.04	2.14
	mit Silbenmethode	9	6.51	10.12	3.37
E zu F2 Kurslevel hoch	ohne Silbenmethode	11	.63	8.97	2.70
	mit Silbenmethode	10	9.66	6.25	1.98

Tab. 62: Gesamtlernerfolg (Einstufung bis Follow-up 2) differenziert nach Kurslevel

Aus Tab. 62 ist zu entnehmen, dass Teilnehmende aus Kursen mit Silbenmethode im Schnitt besser abschneiden als die anderen Teilnehmer, und dies auf allen Kurslevel. Obwohl dieser Zusammenhang überall erkennbar ist, ist er lediglich im levelübergreifenden Vergleich und auf dem höchsten Niveau statistisch signifikant: $t(19) = -2.649, p = .016$.

Da die Silbenmethode in der Projektstudie gesamthaft (unter Berücksichtigung des Umstands, dass dies stark durch die beste Gruppe geprägt ist) einen positiven Einfluss auf das Arbeitsgedächtnis hat, stellt sich die Frage, inwiefern dies für das Abschneiden im A1-Kurs respektive in der telc-Prüfung relevant ist. Eine Korrelation der Ergebnisse bei den Kunstwörtern zu allen drei Messzeitpunkten mit dem telc-Ergebnis lässt keinen oder nur einen schwachen Zusammenhang erkennen: Die Punktezahl bei der Einstufung korreliert nicht signifikant mit dem telc-Ergebnis und das Abschneiden bei den Follow-up-Messungen 1 und 2 korreliert zwar hochsignifikant ($p = .000$ und $p = .000$), aber nur schwach mit dem telc-Resultat ($r = .264$ und $r = .274$). Beim Gesamtlernerfolg zeigt sich keine signifikante Korrelation zwischen Lernerfolg und Kunstwörter (s. Abschn. 6.1.3.3).

Wenn statt des telc-Resultats oder des Gesamtlernerfolgs die einzelnen Fertigkeiten betrachtet werden, so weist das Arbeitsgedächtnis ebenfalls keinen ausgeprägten Zusammenhang auf. Bezüglich der beim telc-Test gemessenen Fertigkeiten korrelieren die Ergebnisse zu den Kunstwörtern bei den Follow-up-Messungen 1 und 2 mit der Fertigkeit Hören. Die Koeffizienten bewegen sich in diesen Fällen jedoch zwischen .263 und .276 ($p \leq .05$), was auf einen schwachen Zusammenhang hindeutet. Hinsichtlich der bei den forschergenerierten Tests erhobenen Leistungen sind die Zusammenhänge jedoch in fast allen Fällen mittel bis stark ausgeprägt. Lediglich die Korrelationen der Kunstwörter mit dem Ergebnis beim ORF im Einstufungstest und beim Hörverstehen in der ersten Follow-up-Messung befinden sich nicht innerhalb des statistischen Schwellenwerts (s. Tab. 63).

		Elicited Imitation	Hörverstehen	ORF	Diktat
Einstufung	Korrelation nach Pearson	.366**	.291*	.249	.476**
Kunstwörter	Signifikanz (2-seitig)	.005	.028	.062	.000
Follow-up 1	Korrelation nach Pearson	.441**	.255	.365**	.509**
Kunstwörter	Signifikanz (2-seitig)	.001	.056	.005	.000
Follow-up 2	Korrelation nach Pearson	.441**	.337**	.372**	.496**
Kunstwörter	Signifikanz (2-seitig)	.001	.010	.004	.000

Tab. 63: Korrelation forschergenerierte Tests differenziert nach Messzeitpunkt und Testverfahren

6.4.1 Überprüfung der Hypothesen

H6. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich des Arbeitsgedächtnisses.

Wie oben beschrieben, bestätigen die Resultate die Hypothese H6 kurslevelübergreifend sowie für das höchste Kurslevel. Das bedeutet, dass die Silbenmethode gesamthaft einen positiven Einfluss auf das Arbeitsgedächtnis hat, wobei sich dieser Effekt besonders deutlich im höchsten Kurslevel zeigt. Für das mittlere und tiefe Kurslevel fällt die Beurteilung der Hypothese hingegen negativ aus.

- H7. Das Arbeitsgedächtnis wirkt sich auf die die vier Fertigkeiten aus.
- H7.1 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Lesen.
 - H7.2 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Schreiben.
 - H7.3 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Sprechen.
 - H7.4 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Hören.

Werden die in den Unterhypothesen H7.1–H7.4 angesprochenen Fertigkeiten anhand der telc-Resultate gemessen, ist kein positiver Zusammenhang mit dem Arbeitsgedächtnis feststellbar (s. Abschn. 6.4), was H7.1–H7.4 falsifiziert, wenn zur Messung der Fertigkeiten das telc-Resultat herangezogen wird. Das Abschneiden der im Rahmen der forschergenerierten Tests gemessenen Fertigkeiten hängt hingegen – mit Ausnahme des Lesens in der Einstufung und des Hörverstehens in der ersten Follow-up-Messung – mit dem Arbeitsgedächtnis zusammen, was sich anhand der mittleren bis starken positiven Korrelation zeigen lässt (s. Abschn. 6.4). Unter der Annahme, dass das Arbeitsgedächtnis die Fertigkeiten direkt beeinflusst und nicht umgekehrt, was zumindest vom theoretischen Standpunkt aus eine plausible Vermutung darstellt (s. Abschn. 4.2.2), validiert dies H7.1–H7.4 weitgehend. Zusammenfassend müssen die Hypothesen dennoch zurückgewiesen werden.

Teil IV: Diskussion

7. Diskussion der empirischen Ergebnisse

Die vorliegende Forschungsarbeit untersucht Zweitschriftlernende und eigens für sie konzipierte Kurse nach Vorgaben des *Konzepts für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* (s. Abschn. 2.1). Aufgrund ihrer Vorkenntnisse in einem nichtlateinischen Schriftsystem stellen Zweitschriftlernende in Alphabetisierungskursen einen Sonderfall dar. Bis anhin ist das Thema in Rahmencurricula und in der Forschungsliteratur kaum präsent.

Diese Forschungsarbeit hat sich daher zum Ziel gesetzt, erstmals Kompetenzen von Zweitschriftlernenden in der Schweiz empirisch zu erfassen und die Effektivität spezifisch auf ihre Bedürfnisse angepasster Kurse zu messen. Dazu wurden zwei Forschungsfragen und sieben Hypothesen aufgestellt (s. Abschn. 3.5 und 3.6), die in den nachfolgenden Abschnitten diskutiert werden.

7.1 Niveau A1 innerhalb von 300 Kurslektionen (Forschungsfrage 1)

Die erste Forschungsfrage der Dissertation lautet:

- F1. Erreichen Zweitschriftlernende, die in der deutschsprachigen Schweiz nach dem Deutschen Zweitschriftlernerkonzept unterrichtet werden, innerhalb von 300 Kurslektionen ein A1-Niveau nach telc?

Insgesamt 19 von 57 Lernenden, was 33 % entspricht, haben die telc-Prüfung A1 nach 300 Kurslektionen erfolgreich absolviert. Der Mittelwert für alle Teilnehmenden liegt bei 54 % der möglichen Punkte und ist somit nur knapp unter der 60 %-Grenze anzusiedeln, die als pass/fail-Schwelle fungiert. Die Erfolgsquote kann zwischen den Institutionen weiter differenziert werden: So haben bei der Institution P fünf (19 %) und bei der Institution Q 14 Teilnehmende (45 %) die telc-Prüfung A1 bestanden. Es ist ein deutlicher Unterschied zwischen den Institutionen zu beobachten, der auch statistisch signifikant ist (t-Test und Regressionsanalyse).

Wenn nur die Erfolgsquote der Institution Q betrachtet wird – mögliche Gründe hierfür werden in Abschnitt 8.2 erläutert –, dann besteht dort eine 45 %-Chance für Zweitschriftlernende, die telc-Prüfung A1 nach 300 Kurslektionen erfolgreich zu bestehen, was für Zweitschriftlernerkurse vielversprechend ist. Diese Werte weisen auf eine deutlich höhere Erfolgsquote als in der Studie von Kurvers hin, in der lediglich «[a] few» Lernende das A1-Niveau nach 300 Kurslektionen erfolgreich absolviert haben (vgl. Kurvers 2015, S. 71).

Hypothese H1 entspricht einer Ausdifferenzierung der Forschungsfrage F1 im Zuge der Anwendung der Unterrichtsmethode:

- H1. Mindestens die Hälfte der Kursteilnehmenden der Treatmentgruppe 1 und 2 erreichen das Niveau A1 nach telc innerhalb von 300 Kurslektionen.

Diese Hypothese untersucht, ob die Studienteilnehmenden das Niveau A1 innerhalb von 300 Kurslektionen erreichen und ob es eine Unterscheidung zwischen beiden Treatments gibt. Von den insgesamt 57 Kursteilnehmenden wurden 30 Lernende mit zusätzlichem Fokus auf der Silbenmethode (Treatmentgruppe 2) und 27 Teilnehmende ohne diesen zusätzlichen Fokus (Treatmentgruppe 1) unterrichtet. Von den insgesamt 19 Personen, welche die telc-Prüfung erfolgreich absolviert haben, stammen zehn Lernende aus der Treatmentgruppe 2, somit hat je exakt ein Drittel der Teilnehmer beider Treatments die Prüfung bestanden. Entsprechend ist der Zusammenhang zwischen Erreichen des A1-Niveaus und der Unterrichtsmethode statistisch nicht signifikant. Auch hinsichtlich der bei der telc-Prüfung erreichten Punktzahl ist durch die t-Tests kein statistischer Zusammenhang nachweisbar, zumindest nicht im kurslevelübergreifenden Vergleich. Auf der Ebene der einzelnen Kurse hingegen unterscheiden sich die Ergebnisse im tiefsten Kurslevel signifikant: Die Gruppe mit der Silbenmethode weist dort einen deutlich höheren Mittelwert auf.

In der Regressionsanalyse existiert jedoch auch im levelübergreifenden Vergleich ein signifikanter Zusammenhang zwischen der Silbenmethode und dem telc-Resultat. Lernende, die mit genannter Methode unterrichtet wurden, weisen in dieser Modellberechnung in der telc-Prüfung A1 ein um 5 Prozentpunkte besseres Ergebnis auf. Bei Betrachtung der Silbenmethode unter Konstanthaltung weiterer Faktoren, wie der L1, dem Alter und dem Kurslevel, tritt ein zwar schwacher, aber positiver Effekt auf die Niveauerreichung A1 zutage. Insgesamt lässt sich festhalten, dass der zusätzliche Unterricht mit der Silbenmethode für die telc-Prüfung von einer gewissen Relevanz ist, da die Silbenmethode einen positiven, wenn auch geringen Einfluss auf die erreichte Punktzahl im telc-Test hat. Besonders für das tiefe Niveau scheint sich die Silbenmethode als vielversprechend zu erweisen.

7.2 Silbenmethode (Forschungsfrage 2)

Die zweite dieser Arbeit zugrunde liegende Forschungsfrage lautet:

- F2. Welche Unterschiede sind in verschiedenen Teilbereichen der Sprachkompetenz feststellbar, wenn Zweitschriftlernende im Rahmen des Zweitschriftlernerkonzepts aus Deutschland mit oder ohne speziellem Fokus auf der Silbenmethode unterrichtet werden?

Um diese Forschungsfrage beantworten zu können, werden die Unterschiede im Gesamtlernerfolg von der Einstufung bis und mit der Follow-up-2-Messung und das

Abschneiden bei den durch den telc-Test gemessenen *Fertigkeiten* betrachtet. Die Unterscheidung der vier Fertigkeiten entspricht den vier Unterhypothesen 2 und 3:

- H2. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im telc-Resultat.
 - H2.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.
 - H2.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.
 - H2.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen.
 - H2.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.
- H3. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich im Lernerfolg.
 - H3.1 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Lesen.
 - H3.2 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Schreiben.
 - H3.3 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Sprechen.
 - H3.4 Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich der Fertigkeit Hören.

Zum besseren Verständnis werden hier nochmals die Kurseinteilung und die Kursphasen dargestellt (s. Tab. 64 und Tab. 65):

Kurslevel	Institution P	Institution Q
hoch	Gruppe P1, mit Silbenmethode	Gruppe Q1, <i>ohne</i> Silbenmethode
mittel	Gruppe P2, <i>ohne</i> Silbenmethode	Gruppe Q2, mit Silbenmethode
tief	Gruppe P3, <i>ohne</i> Silbenmethode	Gruppe Q2, mit Silbenmethode

Tab. 64: Kurslevel und Treatment

Kursphase	Zeitspanne	Messzeitpunkte
eins	Kurslektion 1 bis 100	Einstufung; Follow-up 1
zwei	Kurslektion 101 bis 200	Follow-up 1; Follow-up 2
gesamt	Kurslektion 0 bis 200	Einstufung; Follow-up 2

Tab. 65: Kursphasen und Messzeitpunkte

Werden die Ergebnisse der telc-Prüfung zur Fertigkeit **Sprechen** betrachtet, fällt auf, dass die Silbenmethode nur im mittleren Kurslevel signifikant bessere Ergebnisse als das Pendant ohne Silbenmethode aufweist. Wenn die Fertigkeit anhand des Elicited Imitation gemessen wird, zeigt der Gesamtlernerfolg von der Einstufung bis zum Follow-up 2, dass das hohe Kurslevel mit Verwendung der Silbenmethode signifikant schlechtere Ergebnisse erzielt hat. Dieser statistisch nur teilweise signifikante und partiell gegensätzliche Zusammenhang zwischen dem Einsatz der Silbenmethode und dem Abschneiden bei der Fertigkeit Sprechen überrascht angesichts der Theorie. Da für einen natürlichen Sprechfluss die silbische Struktur von Wörtern bekannt sein muss, sollte die Silbe einen entscheidenden Einfluss auf das Sprechen haben. Entsprechend wird von einigen Autoren postuliert, dass die silbische Struktur von Wörtern den Sprechfluss (vgl. Schröder-Lenzen 2013, S. 33–38; Dahmen/Weth 2018, S. 126–

127) und die Artikulation (vgl. Albert et al. 2015, S. 48–49) optimieren kann. Das Ergebnis der Elicited Imitation auf dem hohen Kurslevel kann diese Annahme nicht stützen, vielmehr weisen die Befunde darauf hin, dass die Silbenmethode einen negativen Einfluss auf das Sprechen hat. Hier könnte argumentiert werden, dass Teilnehmende aus dem hohen Kurslevel bereits Kenntnisse und/oder Kompetenzen mitbringen und daher das Training der Silbenstruktur eher zu Verwirrungen führt. Ein solcher Zusammenhang scheint möglich, lässt sich aber aus der vorliegenden Datenkonstellation nicht erhärten und müsste weiter untersucht werden, etwa mit introspektiven Methoden und/oder Befragungen der Lernenden. Der positive Einfluss der Silbenmethode auf das telc-Resultat in der mittleren Kursgruppe entspricht eher der durch die Theorie gestützten Erwartungshaltung. Da aber beim telc-Resultat und dem Gesamtlernerfolg nur je eines von drei Kurslevel statistisch signifikante Zusammenhänge aufweist und somit zwei Drittel der Zusammenhänge unklar bleiben, bleibt offen, inwiefern genau die Silbenmethode sich positiv auf die Fertigkeit Sprechen auswirkt. Beim **Hörverstehen** verzeichnet der Einsatz der Silbenmethode lediglich für das tiefe Kurslevel signifikant bessere Ergebnisse in der telc-Prüfung. Auch in den selbst generierten Tests weist das tiefe Kurslevel signifikant höhere Werte auf, wenn mit der Silbenmethode unterrichtet wird, und zwar für die erste Kursphase sowie für die Zeitspanne zwischen der Einstufung und dem Follow-up 2. Das Gegenteil gilt für das mittlere Kurslevel. Dort fällt die Lernprogression in denselben Zeitspannen geringer aus, wenn die Silbenmethode eingesetzt wird. Kurz gesagt sprechen die Ergebnisse im tiefen Kurslevel eher für und im mittleren Kurslevel eher gegen die Silbenmethode. In der Theorie wird davon ausgegangen, dass mit der Silbenmethode die auditive Wahrnehmung geschult wird, die für das Hörverständnis notwendig ist. Durch das Training mit der Silbenmethode können die Lernenden, so die Annahmen, bewusster zuhören und dementsprechend die Informationen besser heraushören. Die Ergebnisse auf dem tiefsten Kursniveau bezüglich Gesamtlernerfolg und telc-Resultat bestätigen diese Annahme. Im mittleren Kurslevel hingegen scheint die Silbenmethode eher hinderlich zu sein, denn dort schneidet der Kurs ohne zusätzlichem Fokus auf der Silbenmethode beim forschergenerierten Test zum Hörverstehen signifikant besser ab. Es ist denkbar, analog zu den Ausführungen zur Hypothese H1, (s. Abschn. 7.1) dass die Silbenmethode nur für tiefere Niveaus gewinnbringend ist.

Hinsichtlich der Fertigkeiten **Lesen** und **Schreiben** können bezüglich der Silbenmethode weder bei der A1-Prüfung noch bei den forschergenerierten Tests (ORF, Diktat) statistisch signifikante Unterschiede festgestellt werden.

Zusammenfassend werden die Unterhypothesen H2.1 und H3.1 bezüglich der Fertigkeit Lesen und H2.2 und H3.2 bezüglich der Fertigkeit Schreiben durch die vorliegenden Daten falsifiziert, da keine signifikanten Zusammenhänge erkennbar sind.

Hinsichtlich der Fertigkeit Sprechen kann H2.3 einzig für das mittlere Kurslevel validiert werden, dort schnitt die Gruppe, die mit der Silbenmethode unterrichtet wurde, bei der telc-Prüfung besser ab als ihr Pendant. Was Hypothese H3.3 angeht, so wirkt sich der Einsatz der Silbenmethode auf hohem Kursniveau sogar negativ auf den Gesamtlernerfolg aus. In der formulierten Pauschalität müssen die Unterhypothesen H2.3 und H3.3 somit abgelehnt werden.

Die Unterhypothesen H2.4 und H3.4, die den Einfluss des Unterrichts mit der Silbenmethode auf die Fertigkeit Hören untersuchen, müssen für das höchste Kurslevel falsifiziert werden, da keine Zusammenhänge erkennbar sind. Auf mittlerem Niveau scheint sich das Treatment der Silbenmethode für den Gesamtlernerfolg negativ zu äussern, wohingegen sich für das tiefe Kurslevel bezüglich A1-Prüfung und Gesamtlernerfolg eine positive Auswirkung dieses didaktischen Ansatzes feststellen lässt. Dieser positive Zusammenhang zwischen Silbenmethode und Hörverstehen in der Kursgruppe auf tieferem Niveau ist insofern aufschlussreich, als dies, soweit der Forscherin bekannt, zum ersten Mal empirisch belegt wird. In ihrer Gesamtheit sind die Unterhypothesen dennoch abzulehnen.

In den Bereichen Lesen und vor allem Schreiben sollten dringend weitere empirische Untersuchungen durchgeführt werden, zumal in Anbetracht des Umstands, dass Schreiben bis heute in dieser Zielgruppe, aber auch bei Analphabeten, generell die am wenigsten erforschte Fertigkeit darstellt (vgl. Gillespie 2001, S. 91).

7.3 Diskussion der übrigen Hypothesen

In diesem Abschnitt werden die übrigen Hypothesen, H4 bis H7, diskutiert, angefangen mit der Hypothese H4 hinsichtlich der Schulerfahrung, anschliessend H5, die die L1-Schreibkompetenz thematisiert und schliesslich die Hypothesen H6 und H7, die auf das Arbeitsgedächtnis bezogen sind.

7.3.1 Schulerfahrung

Zur Beantwortung der vierten Hypothese wird die Anzahl der Schuljahre der Teilnehmenden betrachtet:

- H4. Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat und den Gesamtlernerfolg.
 - H4.1 Die Anzahl der Schuljahre hat einen Einfluss auf das telc-Resultat.
 - H4.2 Die Anzahl der Schuljahre hat einen Einfluss auf den Gesamtlernerfolg.

Die Schulbildung der Teilnehmenden wurde anhand folgender Cluster ausgewertet: 1 bis 3 Schuljahre ($n = 5$), 4 bis 6 Schuljahre ($n = 13$), 7 bis 9 Schuljahre ($n = 29$) und 10 bis 12 Schuljahre ($n = 9$). Eine Person gab an, keine Schule besucht zu haben. Auf diesen Fall wird nicht näher eingegangen, weil Einzelfälle in Clusteranalysen nicht berücksichtigt werden können.

Je höher die in Schuljahren gemessene Schulbildung der Teilnehmenden, desto höher ist die Wahrscheinlichkeit, die **telc-Prüfung** erfolgreich zu bestehen. So steigt der Prozentwert mit den Schuljahren an: 20 % für das erste Cluster, 31 % für das zweite, 35 % für das dritte und 44 % für das vierte. Die entsprechende Varianzanalyse weist jedoch keinen statistischen Zusammenhang auf. In der Regressionsanalyse ist hingegen der Zusammenhang zwischen Anzahl Schuljahren und telc-Prüfung für gewisse Cluster signifikant: Lernende mit 4 bis 6 Schuljahren erzielten um 8.5 Prozentpunkte schlechtere Ergebnisse als Lernende mit 7 bis 9 Schuljahren, die die Referenzkategorie darstellen. Lernende mit 10 bis 12 Schuljahren zeigen andererseits um 8.4 Prozentpunkte bessere Ergebnisse als die Referenzkategorie.

In Bezug auf den **Gesamtlernerfolg** sind in der linearen Regression und im direkten Vergleich der Cluster anhand des Mittelwerts keine Zusammenhänge zwischen Schulerfahrung und Lernerfolg ersichtlich.

Die Ergebnisse deuten darauf hin, dass die Anzahl der Schuljahre als Prädiktor für das erfolgreiche Abschneiden beim A1-Test von telc gelten kann. Dieser Befund stützt den in Theorie und Praxis häufig anzutreffende Fokus auf die Anzahl der Schuljahre zur Einstufung der Teilnehmenden (s. Abschn. 1.3.2). Der Umstand, dass der Zusammenhang nur im Regressionsmodell, unter Berücksichtigung weiterer Variablen, nicht aber im direkten Vergleich der Alterscluster mittels t-Test signifikant ist, weist darauf hin, dass die Anzahl der Schuljahre nicht als einziges Kriterium herangezogen werden sollte. So ergibt sich beispielsweise für das Geschlecht ein ähnlich hoher und ebenfalls signifikanter Regressionskoeffizient, obwohl dieser Faktor in Theorie und Praxis selten als Einteilungskriterium gewählt wird.

7.3.2 Schreibkompetenz in der L1

Im Folgenden wird die Schreibkompetenz in der Erstschriftsprache der Lernenden diskutiert:

- H5. Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das Resultat im telc-Test und den Gesamtlernerfolg.
 - H5.1 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst das telc-Resultat.
 - H5.2 Der Grad der Schreibkompetenz in der Erstschriftsprache beeinflusst den Gesamtlernerfolg.

Zur Messung der Schreibkompetenz in der Erstschriftsprache wurden der Wortschatz und die formale Richtigkeit in den von den Teilnehmenden verfassten Kurztexten im vor dem Einstufungstest ausgefüllten Fragebogen herangezogen. Wie sich zeigt, ist der Zusammenhang zwischen der so gemessenen Schreibkompetenz in der Erstschriftsprache und dem Abschneiden in der telc-Prüfung oder dem durch die selbst konzipierten Tests erfassten Gesamtlernerfolg statistisch nicht signifikant, und dies unabhängig vom verwendeten statistischen Testverfahren.

In Abschnitt 1.2.2 wurde postuliert, dass Zweitschriftlernende möglicherweise Wissen aus der Erstschriftsprache in die Zweitsprache transferieren können, analog zu den Befunden aus Studien, die mit anderen Zielgruppen durchgeführt wurden (vgl. Schoonen et al. 2003, S. 166; Leki et al., S. 106). Ein entsprechender Transfer lässt sich aus den Ergebnissen dieser Studie nicht herauslesen, d. h. es ist kein statistisch signifikanter Zusammenhang zwischen Wortschatz und formaler Richtigkeit in der Erstschriftsprache und dem Abschneiden beim Abschlusstest oder dem Gesamtlernerfolg erkennbar.

In diesem Bereich müssten weitere Untersuchungen durchgeführt werden, um Zusammenhänge zwischen Erstschriftsprache und Zweitsprache genauer abzuklären, vor allem, was den häufig postulierten Nutzen des Rückgriffs auf die Muttersprache beim Zweitspracherwerb betrifft (vgl. Heyn 2013; Albert et al. 2015, S. 100–117; Feick/Schramm 2016, S. 217).

7.3.3 Arbeitsgedächtnis

Die Hypothesen H6 und H7 beziehen sich auf das Arbeitsgedächtnis und lauten folgendermassen:

- H6. Treatmentgruppe 1 und Treatmentgruppe 2 unterscheiden sich bezüglich des Arbeitsgedächtnisses.
- H7. Das Arbeitsgedächtnis wirkt sich auf alle vier Fertigkeiten aus.
 - H7.1 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Lesen.
 - H7.2 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Schreiben.
 - H7.3 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Sprechen.
 - H7.4 Das Arbeitsgedächtnis hat einen Einfluss auf die Fertigkeit Hören.

Die Ergebnisse der Kurslevel mit und ohne Silbenmethode unterscheiden sich voneinander dahingehend, dass die Kurse mit zusätzlichem Fokus auf der Silbenmethode auf allen Niveaus

deutlich bessere Resultate im Mottier-Test erzielen. Diese Unterschiede lassen sich statistisch jedoch nur levelübergreifend sowie im hohen Kurslevel belegen. Dies führt zu einer grundsätzlichen Bestätigung der Hypothese H6 und Bejahung der Frage, ob sich die Silbenmethode positiv auf das Arbeitsgedächtnis auswirkt. Wie jedoch die Beantwortung der Unterhypothesen zeigt, sind diese Befunde unter Vorbehalt zu betrachten.

Die Hypothesen H7.1–H7.4, die untersuchen, inwiefern das Arbeitsgedächtnis und die vier Fertigkeiten zusammenhängen, bedürfen einer ausführlicheren Erläuterung. Das Abschneiden beim Mottier-Test weist einen signifikanten mittleren bis starken Zusammenhang mit den vier Fertigkeiten auf, der mittels forschergenerierter Tests gemessen wurde. Die vier Fertigkeiten, wie sie im telc-Test untersucht werden, hängen hingegen nicht oder nur schwach mit dem Resultat bei den Kunstwörtern zusammen. Je nach Messart der vier Fertigkeiten werden somit unterschiedliche Ergebnisse erzielt.

Wenn nun das Abschneiden der Teilnehmenden bei den Kunstwörtern zu allen drei Zeitpunkten als unabhängige Variable in die lineare Regression aus dem Abschnitt 6.2.3 mit dem telc-Resultat als zu erklärende Variable mit hineingenommen wird, zeigt sich, dass die Regressionskoeffizienten der Kunstwörter statistisch nicht signifikant sind. Dies stützt den Befund aus den Korrelationen, wonach das Arbeitsgedächtnis für das Absolvieren der telc-A1-Prüfung nicht relevant ist.

Wird das lineare Regressionsmodell aus Abschnitt 6.3.4 verwendet, wo der Lernerfolg bis zum zweiten Messzeitpunkt hin gemessen wird, weist das Ergebnis des Mottier-Tests bei der Einstufung (nicht aber bei den Follow-up-Messungen) einen signifikanten Regressionskoeffizienten auf ($B = .34$, $p = .013$). Das Arbeitsgedächtnis scheint somit tendenziell einen positiven Einfluss auf das Abschneiden bei den eigens für die Studie entwickelten Tests zu haben, nicht aber auf die telc-Prüfung.

Verschiedene Auslegungen dieser Resultate sind denkbar. Eine mögliche Erklärung ist, dass das Arbeitsgedächtnis zwar einen positiven Einfluss auf den Lernerfolg und die vier Fertigkeiten hat, dies jedoch von der telc A1-Prüfung nicht widerspiegelt wird. Mit anderen Worten, das Arbeitsgedächtnis wäre für die telc-Prüfung von untergeordneter Relevanz, d. h. ein grösserer Arbeitsspeicher brächte beim Absolvieren der A1-Prüfung kaum Vorteile. Da es aber für einen Erfolg beim Hör- und Leseverstehen nötig ist, relevante Informationen im Arbeitsspeicher zu behalten und diese dann abzurufen, scheint die Argumentation, wonach

ein besseres Arbeitsgedächtnis nicht zu besseren Resultaten beim telc-Test führen soll, wenig überzeugend.

Eine weitere Möglichkeit, den niedrigen Korrelationskoeffizienten zwischen Kunstwörtern und den telc-Resultaten zu interpretieren, besteht in der Annahme, dass Kunstwörter kein adäquates Testinstrument sind, um das Arbeitsgedächtnis genau zu erfassen. Dies widerspricht der Ansicht und den Ergebnissen von Castro-Caldas et al. (1998) (s. Abschn. 4.2.2), die Kunstwörter als geeignete Messmethode für die phonologische Bewusstheit und somit den Arbeitsspeicher betrachten. Es stimmt jedoch mit den Überlegungen und Diskussionen aus der Diagnostik und der Therapie überein, die für die Messung des Arbeitsspeichers auf eine Vielzahl unterschiedlicher Verfahren setzen und die Verwendung eines einzelnen Instruments, wie hier der Kunstwörter, für unzureichend halten (vgl. Hasselhorn et al. 2000, S. 127; Böhme/Arnold 2008, S. 163). Eine Konsequenz aus dieser Argumentation könnte sein, dass das Mottier-Verfahren zwar tatsächlich die Kapazität des Arbeitsgedächtnisses misst, das Resultat jedoch durch die Verwendung der Silbenmethode überschätzt wird. Das Testinstrument basiert auf dem Hören einzelner, deutlich artikulierter Silben und dem Nachsprechen derselben. Die Silbenmethode geht in ihrem Grundgedanken auf die gleiche Idee zurück. Das heisst, dass Teilnehmende, die mit Hilfe dieser Lehrmethode geschult wurden, im Umgang mit auf diese spezifische Weise intonierten Silben geschickter sind und daher möglicherweise auch bei den Kunstwörtern besser abschneiden. Dadurch würde ihr Arbeitsspeicher entsprechend überschätzt werden, was die teilweise widersprüchlichen statistischen Zusammenhänge erklären könnte. Der Mottier-Test wäre in der vorliegenden Studie somit eher ein Indikator für die erfolgreiche Implementierung der Silbenmethode und liesse keine Rückschlüsse auf das Arbeitsgedächtnis zu. In diesem Licht betrachtet, wäre die Silbenmethode nur im höchsten Kurslevel wirksam umgesetzt worden und für die oben besprochenen Hypothesen H6 und H7 müsste geschlussfolgert werden, dass keine Aussage getroffen werden kann. Zum Testen des Arbeitsgedächtnisses bei gleichzeitiger Verwendung der Silbenmethode scheint für zukünftige Studien eine andere Messmethode als der Mottier-Test oder zumindest eine ergänzende Verwendung weiterer Testinstrumente sinnvoll.

7.4 Weitere Ergebnisse

Im Rahmen der vorliegenden Untersuchung wurde ein umfangreicher Datensatz zu Zweitschriftlernenden generiert. Aus der statistischen Auswertung dieser Daten sind weitere,

nicht über die Forschungsfragen oder Hypothesen abgedeckte Zusammenhänge ersichtlich geworden. Auf diese soll an dieser Stelle näher eingegangen werden.

So fällt bei Betrachtung der **Kurslevel** auf, dass mehr als ein Sechstel der Streuung der telc-Prozentwerte durch das Kurslevel erklärt werden kann. Das höchste Kurslevel hat bei der telc-Prüfung am besten abgeschnitten; 48 % der Teilnehmenden haben die Prüfung bestanden. Im Vergleich dazu haben 35 % der Teilnehmenden aus dem mittleren und 16 % aus dem tiefen Kurslevel die telc-Prüfung erfolgreich absolviert. Die Wahrscheinlichkeit, die telc-Prüfung zu bestehen, steigt somit, je höher das Kurslevel ist. Dieses Ergebnis mag nicht verwundern, da Teilnehmende im höchsten Kurslevel beim Einstufungstest am besten abschnitten und somit die meisten Vorkenntnisse in den Unterricht mitbrachten. Dieses Vorwissen dürfte dann auch bei der telc-Prüfung dazu geführt haben, dass sie bessere Ergebnisse erzielten als Partizipanten aus den tieferen Kurslevel. Umgekehrt könnten diese Vorkenntnisse zu einer Art Plafonierung des Lernens geführt haben: Dadurch, dass Personen im hohen Kurslevel bereits über mehr Wissen als andere Teilnehmende verfügten und weniger Potenzial für Progression gegeben war, zeigte sich auch ihr Lernfortschritt eingeschränkt. Zumindest legt das die deskriptive Analyse nahe, denn der erzielte Lernfortschritt vom Einstufungstest bis zur telc-Prüfung steigt, je tiefer das Niveau ist. Eine mögliche Implikation dieser Resultate besteht darin, dass Teilnehmende mit hohen bis sehr hohen Vorkenntnissen in der Zweitsprache Deutsch in Kursen mit (noch) schnellerem Lerntempo besser aufgehoben wären. Zudem untermauern die Ergebnisse die Wichtigkeit einer korrekten Einstufung. Nur dadurch erhalten Teilnehmende die Möglichkeit, in für sie geeigneten Kursgruppen mit angemessenem Tempo zu lernen.

In Anbetracht der obigen Ausführungen scheint die Annahme naheliegend, dass Teilnehmende, die bereits vor Beginn der Studie einen **Deutschkurs** besucht haben, aufgrund ihres höheren Vorwissens bei der telc-Prüfung besser abschneiden und einen geringeren Lernfortschritt erzielen. Zumindest Letzteres trifft zu: Personen, die bereits einen Sprachförderkurs besucht haben, weisen einen geringeren durchschnittlichen Gesamtlernerfolg auf. Dieser Umstand kann analog zu den obigen Ausführungen auf die Plafonierung des Lernerfolgs zurückgeführt werden. Auffällig ist aber, dass ein vorgängig besuchter Deutschkurs einen signifikant *negativen* Einfluss auf das telc-Ergebnis hat. Mit einem Regressionskoeffizienten von rund -21 ist dieser Zusammenhang zudem sehr stark ausgeprägt. Die Quote bestandener A1-Prüfungen bei Personen ohne vorgängige

institutionelle Sprachförderung liegt bei fast zwei Dritteln, wohingegen der Anteil bestandener Prüfungen ein Drittel ausmacht für die Gruppe der Teilnehmenden, die bereits Kurse besucht hat. Somit weisen Lernende mit Deutschkurserfahrung nicht nur eine geringere Lernprogression auf, sondern schneiden insgesamt auch merklich schlechter ab als die anderen Teilnehmenden.

Verschiedene Interpretationen scheinen hier plausibel. Eine erste Lesart der Resultate ist, dass der vorgängig besuchte Deutschkurs für Zweitschriftlernende möglicherweise nicht bedürfnis- und bedarfsgerecht ausgerichtet war. So haben die Partizipanten in diesen Kursen unter Umständen Lernstrategien entwickelt, die für sie nicht effektiv waren und für ihre Lernprogression nicht förderlich. Es lässt sich anhand der verfügbaren Daten nicht beurteilen, inwiefern dies zutrifft. Auch möglich scheint, dass die Motivation der Teilnehmenden bei einem erneuten Kursbesuch nachlässt. Personen, die im Vorfeld beispielsweise einen A1- oder gar A2-Kurs besucht haben, könnten das in der Forschungsarbeit benutzte Kurskonzept als Rückstufung empfunden haben (im Sample dieser Arbeit waren Personen vertreten, die einen solchen Kurs besucht hatten, obwohl ihre Lese- und Schriftkenntnisse gemäss Einstufungstest unzureichend waren). Wer bereits an Alphabetisierungskursen teilgenommen hat, könnte hingegen das Kursformat mit vorgängiger Alphabetisierungsphase als unnötige Wiederholung betrachtet haben.

Unabhängig von der gewählten Auslegung wird erneut deutlich, wie entscheidend die korrekte Ersteinstufung für Teilnehmende ist, um den Kurslernenden die für sie passenden Kursgefässe anbieten und unnötige Kurswiederholungen vermeiden zu können. Diese Einstufung obliegt in der deutschsprachigen Schweiz den Sprachkursanbietern, d. h. ob und wie eine Einstufung vorgenommen wird, ist Sache der Kursinstitution. Da nicht auf ein standardisiertes Einstufungsverfahren zurückgegriffen werden kann, geschieht dies häufig anhand von Einstufungstests in Lehrwerken (z. B. *Schritte plus Neu*), die für Analphabeten wie auch Zweitschriftlernende meist ungeeignet sind, da sie den Fokus stark auf Grammatik legen. So ist in der Praxis regelmässig zu beobachten, dass Analphabeten und/oder Zweitschriftlernende erst nach einigen Monaten als im lateinischen Alphabet schriftunerfahren erkannt werden. Ein solches System ist ineffizient: Einerseits kann es die Motivation durch nichtlineare und sprunghafte Kurswechsel mindern und andererseits wird so die ohnehin limitierte Sprachpauschale der Lernenden nicht zielführend eingesetzt.

Es lässt sich somit ein erhebliches Forschungsdesiderat mit hoher Praxisrelevanz ausmachen: Zum einen sollte für die Praxiswelt überlegt werden, wie Teilnehmende standardisiert eingestuft werden können und zum anderen sollten Lehrpersonen adäquat ausgebildet bzw. weitergebildet werden, damit ein schriftsprachliches Defizit bei Teilnehmenden gleich zu Kursbeginn bemerkt werden kann und weitere Schritte eingeleitet werden können.

In der direkten Gegenüberstellung schnitten Frauen und Männer bei der telc-Prüfung ähnlich gut ab, und zwar bezüglich pass/fail-Quote wie auch hinsichtlich der erreichten Punkte. Das **Geschlecht** weist jedoch in der Regressionsanalyse einen signifikanten Einfluss auf das telc-Resultat auf. Demnach schlossen weibliche Teilnehmende unter ansonsten gleichbleibenden Bedingungen die A1-Prüfung um fast acht Prozentpunkte besser als männliche Teilnehmende ab. Dieses Resultat widerlegt einige Zweitspracherwerbsuntersuchungen, die zeigen, dass männliche Lernende einen höheren Lernfortschritt erreichen als weibliche Teilnehmende (vgl. z. B. van Tubergen 2010, S. 528; Scheible/Rother 2017, S. 15). Das Ergebnis ist insofern aufschlussreich, als eine der Kontrollvariablen im Regressionsmodell prüft, ob die Teilnehmenden Kinder haben: Werden Kinderbetreuungspflichten konstant gehalten, schneiden Frauen besser ab als Männer.

Hinsichtlich der **Erstsprache** lassen sich nur Aussagen bezüglich Arabisch, Persisch und Tigrinya sprechenden Teilnehmenden machen. Die übrigen Erstsprachen wiesen in keinem statistischen Testverfahren signifikante Unterschiede auf. Lernende mit Arabischer Erstsprache schnitten am besten ab. Sie bestanden relativ gesehen am häufigsten die telc-Prüfung und in der Regressionsanalyse weisen sie den höchsten Koeffizienten auf; im Vergleich zu Tigrinya als Referenzkategorie erreichten Personen mit Arabisch als L1 in der telc-A1-Prüfung 14 Prozentpunkte mehr. Umgekehrt weisen Persisch sprechende Teilnehmende im Regressionsmodell um fast 18 Prozentpunkte niedrigere telc-Prozentwerte als Personen mit Tigrinya als L1 auf. Diese Ergebnisse spiegeln die relativ hohe Literalitätsquote in Syrien (im Jahr 2012 lag sie bei 84 %) im Vergleich zu Afghanistan (im Jahr 2011 bei 32 %). Eritrea liegt dazwischen mit 69 % für das Jahr 2012 (vgl. UNICEF 2013a; b; The World Bank 2019). Die Daten dieser Untersuchung bestätigen demnach die Literalitätsstatistiken: Je höher die Literalitätsquote des betreffenden Landes ist, desto höher fallen die telc-Prozentwerte aus. Insgesamt lässt sich 17 % der Streuung der telc-Prozentwerte durch die Erstsprache erklären. Da die jeweiligen Schriftsysteme sehr unterschiedlich sind, könnte argumentiert werden, dass eine Trennung der Zweitschriftlernerurse nach Erstsprache resp. Erstschriftsprache sinnvoll

sein könnte. Dadurch könnte auch ein Rückgriff auf die Muttersprache, der von verschiedenen Autoren als begünstigend postuliert wird, einfacher gestaltet werden (vgl. Heyn 2013; Albert et al. 2015, S. 64–71; Feick/Schramm 2016, S. 217). Dennoch erweist sich eine solche Trennung in der Praxis als eher schwierig umsetzbar, zumal diese für die Teilnehmenden auch als hinderlich erscheinen kann, denn Lehrpersonen haben in sprachlich homogenen Kursgruppen häufig Schwierigkeiten, die Teilnehmenden zur Benutzung der Zweitsprache anzuregen. Zudem wäre eine solche Homogenisierung ungünstig hinsichtlich der Entwicklung der interkulturellen Kompetenzen der Teilnehmenden.

Die im Rahmen der Forschungsarbeit erhobenen Daten widersprechen dem Befund von Schuller et al., wonach der Fortschritt in der Zweitsprache bei Lernenden mit **Fremdsprachenkenntnissen** aufgrund bereits verfügbarer linguistischer Lernstrategien höher ausfällt (vgl. Schuller et al. 2011, S. 161). Die vorliegenden Resultate deuten im Gegenteil an, dass Lernende mit Fremdsprachenkenntnissen beim Gesamtlernerfolg schlechtere Ergebnisse aufweisen als monolinguale Teilnehmende. Gründe hierfür könnten negative Transfers aus der Fremdsprache in die Zielsprache sein (s. Abschn. 1.2.2).

Im *Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* wird für die Teilnahme an *Zweitschriftlernerkursen* ein beruflicher Abschluss und/oder eine mehrjährige Schulausbildung verlangt (vgl. BAMF 2018a, S. 10–11). Einen positiven Einfluss der Schulbildung auf das Abschneiden beim telc-Test ist aus der linearen Regression in Abschnitt 6.2.3 ersichtlich. Auffällig ist aber, dass die Berufsbildung und Erfahrung auf dem Arbeitsmarkt nach den vorliegenden Daten für das Abschneiden bei der A1-Prüfung eher hinderlich zu sein scheint. So fällt das telc-Resultat für Personen mit einer **Berufslehre** und/oder **Arbeitserfahrung** im Heimatland geringer aus, was den Aussagen von Scheible und Rother widerspricht, nach denen sich die Arbeitserfahrung nicht auf die Entwicklung der Deutschkenntnisse auswirken soll (vgl. Scheible/Rother 2017, S. 23). Möglicherweise führen die abgeschlossene Lehre oder der Umstand, dass bereits gearbeitet wurde, zu einer Art Unwillen, wieder ‚in die Schule‘ gehen zu müssen, wobei dies nur eine Vermutung ist, die weiter untersucht werden müsste. Daraus ergibt sich jedoch die Frage, inwiefern berufliche Abschlüsse für die Einteilung von Zweitschriftlernenden sinnvoll sind, wenn Personen mit einer entsprechenden Ausbildung signifikant schlechtere Ergebnisse erzielen. Sollten sich diese Resultate in weiteren Studien erhärten, müssten die Bedingungen für die Teilnahme an Zweitschriftlernerkursen aus dem deutschen Konzept überdacht werden.

Weniger überraschend sind die Ergebnisse bei der Betrachtung der Frequenz des **Deutschgebrauchs** im Alltag und deren Einfluss auf die telc-Prüfung. Die entsprechenden Koeffizienten liegen beide im Bereich des statistisch Signifikanten: Bei ein- bis dreimaligem Gebrauch des Deutschen pro Woche werden bei der telc-Prüfung 16 Prozentpunkte mehr erreicht und wer vier- bis siebenmal pro Woche auf Deutsch kommuniziert, erreicht bei ansonsten gleichbleibenden strukturellen Merkmalausprägungen beim Abschlusstest fast 17 Prozentpunkte mehr. Diese Ergebnisse zeigen die wesentliche Rolle, die der Gebrauch der Zielsprache ausserhalb der Kurstrukturen beim Sprachenerwerb spielt, wie dies schon Kurvers oder Scheible und Rother in ihren Studien nachgewiesen haben (vgl. Kurvers 2015, S. 73; Scheible/Rother 2017, S. 23). Gleichzeitig scheint aber die Häufigkeit selbst nicht entscheidend zu sein, d. h. für den Prüfungserfolg reicht es offenbar schon aus, wenn überhaupt ausserhalb des Unterrichts auf Deutsch kommuniziert wird. Dies unterstreicht die Bedeutung ergänzender Sprachförderangebote wie Tandems und Praktika oder generell der Umgang mit deutschsprachigen Personen ausserhalb der Kursstruktur.

8. Limitationen und Generalisierbarkeit der Studie

Im vorliegenden Kapitel werden zuerst die Wahl und die Limitationen der Testinstrumente und des Forschungsdesigns erläutert (Abschn. 8.1). Auf eine Diskussion der möglichen theoretischen Einwände gegen die einzelnen Verfahren wird hier verzichtet, denn diese wurden bereits in Kapitel 4 behandelt; dasselbe gilt für die Testgüte in Abschnitt 6.1. Wie sich gezeigt hat, ist in der vorliegenden Studie die Wahl der externen Forschungspartner von erheblichem Einfluss auf die Resultate, worauf genauer eingegangen wird (s. Abschn. 8.2). Anschliessend erfolgt eine Besprechung der Generalisierbarkeit der im Rahmen dieser Arbeit gewonnen Befunde (s. Abschn. 8.3).

8.1 Testinstrumente und Forschungsdesign

Um die allgemeine Sprachkompetenz der Teilnehmenden zu messen, wurden einerseits forschergenerierte Tests während des Kursverlaufs und andererseits die telc-Prüfung A1 zum Ende des Kurses durchgeführt. Der telc-Test wird, neben dem Goethe-Zertifikat, im *Zweitschriftlernerkonzept* als Abschlusstest empfohlen (vgl. BAMF 2018a, S. 21). Im Kanton Bern hat sich beim Testen des A1-Niveaus in der untersuchten Zielgruppe der telc-Test etabliert, weshalb er auch hier gewählt wurde. Bei der telc-Prüfung handelt es sich um ein professionell entwickeltes und grossflächig empirisch validiertes Testverfahren, das mit der Q-Mark der ALTE ausgezeichnet wurde.³⁶ Für den Einstufungs- und die Zwischentests konnte nicht auf zertifizierte Testverfahren zurückgegriffen werden, da bisher keine derartigen Instrumente vorliegen. Stattdessen wurde eigens für das Pilotprojekt ein Testverfahren kreiert, wobei Testinstrumente gewählt wurden, die bereits mit Zweitschriftlernenden auf Deutsch oder einer anderen Zweitsprache oder zumindest mit Jugendlichen im DaZ-Bereich erfolgreich angewendet wurden (s. Abschn. 4.2).

Die ausgewählten Testinstrumente zeigen in den für diese Qualifizierungsarbeit gewonnen Daten häufig eine signifikante, mittelstarke Korrelation mit den jeweils korrespondierenden Fertigkeiten bei der telc-Abschlussprüfung (s. Abschn. 6.1.3.3). Einzige Ausnahme ist die ORF, die nicht mit dem telc-Test Lesen korreliert, dafür aber mit Sprechen. Bei Hasbrouck und Tindal ist eine starke Korrelation von ORF und Textverstehen vor allem bei Primarschulkindern

³⁶ Die Q-mark ist ein von der ALTE entwickelter Qualitätsstandard, der 17 Indikatoren umfasst, welche Testkonstruktion und -analyse, Testbewertung, Administration und Logistik sowie Kommunikation mit Stakeholdern würdigen (vgl. ALTE o. J.).

und Sekundarschüler zu beobachten, wohingegen der Zusammenhang bei älteren Lernenden kaum vorhanden ist (vgl. Hasbrouck/Tindal 2006, S. 636). Die Teilnehmenden am Pilotprojekt sind 19 Jahre alt und älter, was möglicherweise die mangelnde Korrelation erklärt. Die Studie von Lems konnte hingegen auch für erwachsene Teilnehmende einen positiven Zusammenhang nachweisen (vgl. Lems 2003, S. 116–125), weshalb hier kein Votum für oder wider den endgültigen Ein- oder Ausschluss der ORF für zukünftige Untersuchungen mit der Zielgruppe (oder Erwachsenen allgemein) erfolgen soll. Weitere Forschung zur ORF müsste zeigen, ob und wie stark beispielsweise das Ergebnis von der Wahl des Lesetexts abhängt. Wie erwähnt korrelieren aber, abgesehen von der ORF, die anderen Testinstrumente und das komplette forschergenerierte Testverfahren mit der telc-Prüfung, was die Verwendung beider Testverfahren im gleichen Untersuchungsdesign rechtfertigt.

In der vorliegenden Studie wurde auf qualitative Erhebungsinstrumente wie beispielsweise Lernerinterviews verzichtet. Da es sich hierbei um die erste grössere Studie mit Zweitschriftlernenden handelt, konnte kaum auf bestehendes Material zurückgegriffen werden. Es schien deshalb plausibel, zunächst quantitative Daten zu erheben, die als erste Wegweiser dienen können, bevor weiter qualitativ erforscht wird. Zudem sollte eine zusätzliche Beanspruchung der Teilnehmenden vermieden werden, da das bestehende Testverfahren bereits vergleichsweise umfangreich ist und ca. 30 Minuten für Hören und Schreiben sowie ca. 20 Minuten für das Sprechen, die Kunstwörter und das Lesen in Anspruch nimmt. Ein letzter Grund waren fehlende Ressourcen; um qualitative Daten zu erheben, wären etliche Stunden Dolmetscher-Arbeit nötig gewesen, die im Rahmen der Dissertation nicht zur Verfügung gestellt werden konnten.

Die vorliegende Arbeit versteht sich als quantitatives und nicht als qualitatives Unterfangen: Es handelt sich um eine experimentelle Interventionsstudie, die Angaben zur kausalen Beziehung zwischen *Zweitschriftlernerkonzept* mit/ohne Silbenmethode und dem Abschneiden beim Abschlusstest respektive dem Gesamtlernerfolg machen soll. Entsprechend liefert die Untersuchung Evidenz für Zusammenhänge mit begründendem Charakter, erlaubt es aber nicht, konkrete kognitive Prozesse bei Zweitschriftlernenden aufzudecken. Bezüglich der Ermittlung kausaler Beziehungen scheint ein gewisser Erfolg vorzuliegen, schliesslich konnten in den Regressionsmodellen 80 % (für die telc-Prüfung) und 48 % (für den Gesamtlernfortschritt) der abhängigen Variablen durch die verwendeten unabhängigen Variablen erklärt werden. Dennoch lassen sich konfundierende Variablen nicht

gänzlich ausschliessen: So könnten die Motivation, das Umfeld oder weitere Variablen, die in dieser Studie nicht erhoben wurden, einen unentdeckten Einfluss auf das Resultat haben.

Hinsichtlich des Forschungsdesigns wäre die Entwicklung und anschliessende Pilotierung eines Itempools wünschenswert gewesen, um für den Echteinsatz nur geeignete Items auswählen zu können. Dies war jedoch nicht möglich, weil Zweitschriftlernende in der Schweiz bisher gemeinsam mit Analphabeten in Kursen unterrichtet werden und somit keine geeigneten Kursgefässe für die Pilotierung der Items existierten. Daher wurde auf die Häufigkeitsliste deutscher Wörter von Tschirner (2016) zurückgegriffen, die aus einem über vier Millionen Wörter umfassenden Korpus gewonnen wurde. Dieses Verfahren hat sich als sinnvoll erwiesen, nach den beiden Pilotierungen (bzw. Erprobungen) mussten nur wenige Items ausgetauscht werden. Die für die eingesetzten Tests benutzten Items befanden sich weitgehend in den einschlägigen Wertebereichen (s. Abschn. 6.1), weshalb keine nachträgliche Adjustierung vorgenommen werden musste.

Die Tonaufnahmen wurden alle in einem professionellen Studio und zur gleichen Zeit aufgenommen. Dennoch gelang es nicht immer, jede Tonaufnahme in gleich hoher Ton- und Stimmqualität und Deutlichkeit der Aussprache aufzunehmen. Es ist daher nicht auszuschliessen, dass bestimmte Items von den Teilnehmenden schlechter verstanden wurden als andere. Gleichzeitig gilt dies für alle Teilnehmenden und alle drei Messzeitpunkte gleichermassen, womit die Vergleichbarkeit der Resultate bestehen bleibt.

Bei der Durchführung der Tests zu verschiedenen Messzeitpunkten wurde die Forscherin für den mündlichen Teil jeweils von demselben Forschungsassistentenpool unterstützt. Diese Personen wurden alle vorgängig geschult und hatten sich an Testvorgaben zu halten. Da der gesamte Testablauf mittels Audiodaten dokumentiert wurde, konnte die Einhaltung derselben überprüft werden. Das Diktat und das Hörverstehen wurden jeweils durch die Forscherin selbst durchgeführt. Die Teilnehmenden sassen hierbei getrennt an Einzeltischen und wurden von der Forscherin kontrolliert, ein Abschreiben beim Tischnachbarn lässt sich jedoch nicht gänzlich ausschliessen.

Zur Erstellung der Fragebogen wurde eng mit grösstenteils zertifizierten erstsprachlichen Dolmetschern zusammengearbeitet (s. Abschn. 4.3.2). Es ist dennoch denkbar, dass die Fragen nicht in allen Sprachen im gleichen Wortlaut gestellt wurden und teilweise abweichende Konnotationen enthielten, obwohl die Übersetzenden über einen grossen Erfahrungsschatz verfügen. Diese Schwierigkeit ist nicht zu vermeiden, wenn Fragebögen oder andere

Textdokumente in andere Sprachen übersetzt werden müssen, die der Forschenden nicht oder zu wenig bekannt sind.

Ein von den Teilnehmenden vielfach geäusselter Wunsch war, die korrigierte Fassung der Diktate kurz nach Testdurchführung zurückzuerhalten, damit sie ihre Fehler nachvollziehen und verbessern könnten. Dem konnte leider nicht entsprochen werden. Die ausgefüllten Antwortblätter mussten zwecks Korrektur durch die Forscherin und den zusätzlich herangezogenen Bewertenden zurückbehalten werden. Der Korrekturprozess nahm in manchen Fällen einige Zeit in Anspruch, weil auf die zeitlichen Ressourcen von Freiwilligen Rücksicht genommen wurde. Eine Rückgabe nach der Korrektur wäre zudem aufgrund der Retest-Items ungünstig gewesen. Zwar hätten die Lernenden durch eine Prüfungseinsicht aus ihren Fehlern lernen können und sich somit möglicherweise verbessert, ein Testverfahren mit zeitnaher Rückgabe hätte aber einen Verzicht auf Retest-Items bedingt. Letzteres entspräche eher der Situation in der Praxis. Hier manifestiert sich ein Dilemma, mit dem Forschende häufig konfrontiert sind.

Zuletzt soll noch angemerkt werden, dass die Stichprobe mit $N = 57$ zwar grösser ist als bei vergleichbaren Studien, aber für die teilweise sehr detaillierte Aufschlüsselung der Variablen immer noch (zu) klein ist. Eine grösser angelegte Studie könnte möglicherweise weitere statistisch signifikante Zusammenhänge nachweisen. Dies gilt insbesondere für die lineare Regression mit dem Gesamtlernerfolg als abhängige Variable.

8.2 Forschung mit externen Partnern

In Forschungsarbeiten mit externen Forschungspartnern wird im Normalfall davon ausgegangen, dass diese aufrichtige und wahrheitsgetreue Auskünfte erteilen (vgl. Aguado 2000, S. 126). In der vorliegenden Studie hat sich jedoch nach einigen Kurswochen auf Seiten der Institution P herauskristallisiert, dass diese nicht vollumfänglich hinter dem Projekt respektive der Art und dem Inhalt der Kurse stand und diesbezügliche Meinungsverschiedenheiten auch offen kommunizierte, bis hin zur Weigerung, die Kurse wie geplant und projektiert weiterzuführen. Nach einer Aussprache sowie telefonischer und elektronischer Kommunikation wurde der Unterricht dann weitergeführt, und zwar, gemäss Aussage der Institution, wie abgesprochen. Aufgrund von Ungereimtheiten in den durch die betroffenen Lehrpersonen mündlich gemachten Aussagen zur Semesterplanung und anderslautenden Informationen einiger Kursteilnehmenden, wurden in der Folge mehrere informelle Gespräche mit einzelnen Lernenden mit und ohne Dolmetscher geführt. Die in

diesem Rahmen eingeholten Angaben standen teilweise in starkem Widerspruch zu den von den Lehrpersonen später auch schriftlich abgegebenen Semesterplänen, sei es hinsichtlich der Anzahl behandelter Kapitel oder der behandelten Inhalte. So wurden manche Kapitelvorgaben deutlich nicht erreicht; dies betrifft vor allem das tiefe und das hohe Kurslevel in der Behandlung des Niveaus A1.2 und die Vorbereitung der telc-Prüfung A1. Beim mittleren Kurslevel lässt sich dies ebenfalls feststellen, jedoch weniger ausgeprägt. Was die in den Kursen behandelten Inhalte und Kapitel angeht, wird bewusst auf konkrete Angaben verzichtet, weil dies einen eindeutig quantifizierbaren Kursfortschritt suggerieren würde. In der Praxis ist es vielmehr so, dass auch Inhalte späterer Kapitel behandelt werden können, ohne mit den Teilnehmern explizit das betreffende Kapitel im Buch zu behandeln.

Im Sinne einer neutralen Forscherperspektive muss mit solchen widersprüchlichen Informationen vorsichtig umgegangen werden. Die Vermutung, dass die Vorgaben nicht eingehalten wurden, erhärtet sich jedoch bei Betrachtung der telc-Resultate³⁷ (s. Abschn. 6.2). Bei Institution P hat im tiefen Kurslevel kein Teilnehmer die telc-Prüfung bestanden, im hohen Level sind es halb so viele wie bei Institution Q. Auf dem mittleren Niveau schneidet Institution P ebenfalls schlechter ab, dies jedoch weniger stark ausgeprägt als bei den anderen Kurslevel. Somit ist Institution Q in jedem Kurslevel besser, und dies – zieht man die Angaben der Teilnehmenden als Beleg heran – umso ausgeprägter, je deutlicher die Vorgaben bei Institution P gemäss Angaben der Kursteilnehmenden nicht eingehalten worden sind.

Von Seiten der Institution P wurde das Projekt unter anderem kritisch beurteilt, weil es «unmöglich sei» (Wortlaut einer Einschätzung seitens einer Kursleiterin der Institution P) und lediglich zur Überforderung der Teilnehmenden führe. Wie die vorliegende Studie zeigt, scheint ein Erfolg mit den entwickelten und eingesetzten Konzepten durchaus möglich zu sein: Bei Institution Q haben fast zwei Drittel der Kursgruppe auf höherem Niveau die telc-Prüfung bestanden, und dies trifft sogar für knapp ein Drittel der Gruppe auf tieferem Niveau zu. Über alle Kursgruppen hinweg haben 45 % der Teilnehmer der Institution Q, der Institution, die sich gemäss eigener Angaben und Aussagen der Teilnehmenden an das Kurskonzept gehalten hat, die telc-Prüfung bestanden. Man kann somit davon ausgehen, dass zumindest für fast die Hälfte der Kursteilnehmenden die telc-Prüfung innerhalb von 300 Kurslektionen machbar ist, sofern ein ‚Mitziehen‘ der Kursleiter und das Einhalten kurskonzeptioneller Vorgaben

³⁷ Institution Q schneidet zwar auch beim Lernerfolg durchgehend besser ab, aber die Unterschiede zur Institution P sind statistisch nicht signifikant.

vorausgesetzt werden können. Bereits Aguado (2000, S. 125) erwähnt, wie wichtig eine gelingende Kooperation mit den beteiligten Lehrpersonen für die Durchführung und Wirksamkeit einer Studie ist.

Eine Möglichkeit, die oben beschriebenen Schwierigkeiten in zukünftigen Forschungsvorhaben zu vermeiden, wäre die Implementierung strengerer Kontrollmechanismen, z. B. durch eine systematische Befragung der Kursteilnehmenden über den Kursfortschritt. Ein solches Monitoring ist jedoch ressourcenintensiv und würde dem Aufbau eines Vertrauensverhältnisses zwischen den Forschungspartnern und der Forschenden im Wege stehen. Daher muss im Einzelfall beurteilt werden, inwiefern solche Kontrollen tatsächlich sinnvoll sind.

Sinnvoll scheint in jedem Fall ein frühzeitiger, transparenter Dialog zwischen Forschenden und durchführenden Institutionen und Lehrpersonen. Auf diese Weise sollten sich Ungereimtheiten und falsche Erwartungen beider Seiten schon vor der Durchführung des Projekts erkennen und somit idealerweise vermeiden lassen. Dies ist umso wichtiger in Anbetracht der zu beobachtenden Kluft zwischen Praxis und Forschung: Bekanntlich fehlen Personen aus der Praxis zum Teil die theoretischen und forschungsmethodischen Grundlagen, während manche Forschende den Blick auf die Gegebenheiten in der Praxis vermissen lässt. Hier gilt es, voneinander zu lernen und zu profitieren.

Das vermeintliche Nichteinhalten der Kursvorgaben durch die Kursleitenden ist nur ein Faktor, wenn auch ein durchaus plausibler, der bei der Suche nach Gründen für die Unterschiede zwischen beiden Institutionen eine bedeutende Rolle spielen dürfte. Ein weiterer möglicher Grund könnten die Differenzen in den Bildungshintergründen der Lehrpersonen darstellen. Damit Lehrpersonen einen *Zweitschriftlernerkurs* in Deutschland durchführen dürfen, benötigen sie eine Zulassung für Integrationskurse: Sie müssen eine DaF/DaZ-Ausbildung vorweisen oder eine Zusatzqualifizierung abgeschlossen haben (vgl. BAMF 2018a, S. 28). Auch im Rahmencurriculum in der Schweiz wird festgehalten, dass die Sicherung der Qualität der Angebote u. a. an die Qualifikationen der Unterrichtenden gekoppelt ist (vgl. Lenz et al. 2009, S. 62–64).

Bei Institution Q müsste eine von sechs Lehrpersonen eine unverkürzte Zusatzqualifizierung gemäss BAMF abschliessen, während bei Institution P drei der sechs Lehrpersonen eine unverkürzte Qualifizierung absolvieren müssten (s. Abschn. 3.3.3). Somit entsprechen die Kursleitenden in der Institution Q eher den Vorgaben des BAMF als die Lehrpersonen aus

Institution P, was ebenfalls ein Grund für das durchwegs bessere Abschneiden der Kursgruppen bei Institution Q sein könnte. Die mangelnde Ausbildung gemäss den Vorgaben des BAMF und die möglicherweise unzureichende Berücksichtigung der Kursinhalte liegen beide bei derselben Institution vor. Daher lassen sich diese beiden Variablen in den vorliegenden Daten nicht trennen, weshalb keine abschliessende Aussage getroffen werden kann.

Schliesslich und grundsätzlich kann nicht ausgeschlossen werden, dass die Differenz zwischen den Institutionen durch weitere intervenierende und hier nicht gemessene Variablen zustande kam. Auch in diesem Falle behielte aber die Aussage Gültigkeit, dass die vom *Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs)* postulierten Ziele auch im hier untersuchten Kontext realistisch sein können.

8.3 Generalisierbarkeit

In dieser Studie geht es um den kausalen Zusammenhang zwischen dem Unterrichten nach dem *Zweitschriftlernerkonzept* mit oder ohne Fokus auf der Silbenmethode und dem erzielten Gesamtlernerfolg respektive dem Abschneiden bei der A1 telc-Prüfung. Das Forschungsprojekt wurde im deutschsprachigen Kanton Bern bei zwei Institutionen durchgeführt. Dabei stellt sich die Frage, inwieweit die Ergebnisse für den ganzen Kanton, die deutschsprachige Schweiz oder sogar darüber hinaus (etwa die gesamte Schweiz oder das deutschsprachige Europa) Geltung beanspruchen können.

Um zu rekapitulieren: Zur Silbenmethode wurde gezeigt, dass sie einen positiven, aber nur schwachen Einfluss auf die telc-Ergebnisse hat. Bezüglich des Erreichens der Kursziele gemäss BAMF lässt sich festhalten, dass 33 % der Kursteilnehmenden das Kursziel erreichen, respektive 45 % in dem Fall, in dem von der Einhaltung der Konzeptvorgaben ausgegangen werden kann. Während Letzteres als Idealfall bezeichnet werden kann, entspricht Ersteres womöglich eher den Gegebenheiten in der Praxis. Zwischen einem Drittel und der Hälfte aller Kursteilnehmenden erreichen demnach das Niveau A1 innerhalb der angestrebten 300 Kurslektionen.

Da es sich beim Forschungsprojekt um ein Feld- und nicht um ein Laborexperiment handelt, ist eine Verallgemeinerbarkeit der Studie über die Stichprobe hinaus möglich. Durch dieses Forschungsdesign wurde die Studie im gewohnten Lernfeld der Teilnehmenden durchgeführt respektive in einem Umfeld, wie es auch in der Praxis anzutreffen ist. In Anbetracht des Umstands, dass die strukturellen und bildungspolitischen Voraussetzungen für andere

Sprachförderinstitutionen im Kanton Bern dieselben sind, scheint deshalb eine Generalisierbarkeit der Resultate auf Kantonsebene möglich – unter dem Vorbehalt, dass durchaus Unterschiede zwischen einzelnen Institutionen bestehen können. Analoges könnte für die Übertragbarkeit auf die gesamte deutschsprachige Schweiz gelten, d. h. auch in anderen Kantonen sollte das A1-Niveau innert 300 Lektionen für eine gewisse Zahl an Zweitschriftlernenden zu erreichen sein und die Silbenmethode einen schwachen, aber positiven Einfluss auf die Niveauerreichung A1 haben.

Eine Generalisierung auf andere deutschsprachige Länder wäre genauer zu prüfen, da sich u. a. die ethnische Zusammensetzung der Flüchtlinge von der Situation in der Schweiz unterscheidet. Wie sich in der vorliegenden Studie zeigt, hat diese einen starken Einfluss auf das Erreichen der Kursziele. Während in der Schweiz eritreische Flüchtlinge den Grossteil der Kursteilnehmenden ausmachen (vgl. Staatssekretariat für Migration SEM 2019a, S. 9), befinden sich beispielsweise in Deutschland vor allem Personen aus Syrien in dieser Situation (vgl. BAMF 2019c, S. 23). Da Letztere gemäss den vorliegenden Daten besser abschneiden als die eritreischen Kursteilnehmenden, scheint es plausibel, dass in Deutschland die Kursziele noch deutlicher erreicht werden. Das Gegenteil gilt für Österreich, wo viele Flüchtlinge aus Afghanistan stammen (vgl. Bundesministerium Inneres 2018, S. 11–16). Da aber im Sample der Projektstudie nur drei Personen aus Afghanistan vertreten sind, ist bei dieser Generalisierung Vorsicht geboten. Auch bezüglich der Anzahl Teilnehmenden in den Kursen unterscheidet sich die Schweiz von den deutschen Nachbarländern. Unterrichtsklassen umfassen in Österreich bis zu 15 (vgl. Fritz et al. 2006, S. 51) und in Deutschland bis zu 25 Teilnehmende (vgl. BAMF 2019a, S. 6). Die Kursgruppen im Forschungsprojekt bestanden hingegen aus maximal 12 Personen, was den schweizerischen Gegebenheiten entspricht. Gemäss den Befunden von Hattie hat die Klassengrösse aber nur einen kleinen Effekt auf die Lernerleistung, auch bei einer Reduktion der Klassengrösse von 25 auf 15 (vgl. Hattie 2013), wobei die Meta-Studie von Hattie wohlgermerkt nur Befunde zu Kindern und Jugendlichen zusammenfasst.

Inwiefern eine Generalisierbarkeit auf den nichtdeutschen Sprachraum (inkl. frankophone und italophone Schweiz) möglich ist, kann hier nicht näher ergründet werden, da sich das Sprachenlernen je nach Ausgangs- und Zielsprachen mehr oder weniger stark unterscheiden kann. Beim Französischen handelt es sich beispielsweise um eine nichttransparente Sprache (vgl. Buetler et al. 2014, S. 2). Italienisch ist zwar eine transparente Sprache, weist aber eine

andere Silbenstruktur auf als das Deutsche (vgl. Cossu et al. 1995, S. 10). Dies beeinflusst vermutlich die Effektivität der Silbenmethode. Es ist zudem anzunehmen, dass sich Sprachen bezüglich relativer Schwierigkeit für die Lernenden unterscheiden, was die Kurszielvorgaben des deutschen *Zweitschriftlernerkonzepts* je nach Sprache mehr oder minder plausibel erscheinen lassen.

9. Implikationen und Ausblick

In diesem Kapitel soll eine etwas breitere Perspektive eingenommen werden und es sollen mögliche Implikationen des Dissertationsprojekts für die Forschung (s. Abschn. 9.1), die Bildungspolitik (s. Abschn. 9.2) und die Praxis (s. Abschn. 9.3) besprochen werden.

9.1 Forschung

Der Grundgedanke des Forschungsprojekts bestand darin, die Eignung des *Zweitschriftlernerkonzepts* für die Schweiz und jene der Silbenmethode für Zweitschriftlernende zu überprüfen. Gleichzeitig sollten weitere Daten generiert werden, die einem besseren Verständnis der Zielgruppe dienen und einen Denkanstoss für zukünftige Forschung geben.

Wie gezeigt werden konnte, erreicht zwischen einem Drittel und fast der Hälfte der Kursteilnehmenden das Niveau A1 in der vorgegebenen Kurszeit. Im untersuchten Pilotprojekt wurde die Kurszusammensetzung durch die Forscherin gesteuert, im Einzelnen wurden Kursgruppen nach Niveau gebildet, was zu verhältnismässig homogenen Kursgruppen geführt hat. In der Praxis dagegen wird vielfach weniger konsequent oder gar nicht nach den Vorkenntnissen der Kursteilnehmenden getrennt, was zu teilweise sehr heterogenen Gruppen führen kann. Es könnte aufschlussreich sein, zu sehen, inwiefern dies die Zielerreichung positiv oder negativ beeinflusst.

Die zweite Forschungsfrage zur Untersuchung der Effektivität der Silbenmethode beim Erwerb der vier Sprachfertigkeiten kann im Rahmen der Dissertation nur teilweise beantwortet werden. Die Silbenmethode geht mit positiven Ergebnissen im Hörverstehen einher und scheint vor allem für das tiefe Kursniveau innerhalb der ersten 200 Kurslektionen erfolversprechend. Im Sprechen hingegen waren negative Resultate im hohen Kurslevel zu verzeichnen; hinsichtlich des Schreibens und des Lesens bleibt die Wirksamkeit dieser Methode offen. Die starke Koppelung der Variablen *Institution* und *Silbenmethode* sind mögliche Gründe hierfür. Eine komplette Entkoppelung dieser beiden Variablen, beispielsweise indem alle Kurse in derselben Institution durchgeführt werden, könnte diesbezüglich mehr Klarheit schaffen.³⁸

³⁸ Eine entsprechende Folgestudie im Kanton Bern wurde im Frühling 2019 durchgeführt, siehe Guerrero Calle (im Erscheinen).

Weiterer Forschungsbedarf ergibt sich auch jenseits der beiden Forschungsfragen. So konnte in dieser Studie der von anderen Forschenden (vgl. z. B. Castro-Caldas et al. 1998) postulierte positive Einfluss des Arbeitsgedächtnisses auf den Zweitspracherwerb nicht bestätigt werden. Allerdings scheint die Messung des Arbeitsspeichers (s. Abschn. 7.3.3) mit Hilfe des Mottier-Tests bei gleichzeitiger Verwendung der Silbenmethode problematisch. Hier wäre ein Einsatz anderer Instrumente vorzuziehen.

Wie die durch das Forschungsprojekt generierten Daten nahelegen, gilt es zudem, die Definition von Zweitschriftlernenden zu überdenken. So wird die Kategorie der Zweitschriftlernenden, wie in den Grundlagen (s. Abschn. 1.3) ausgeführt, in der Regel nicht weiter differenziert. Vielmehr werden Personen ohne Kenntnisse des Alphabets der Zielsprache, die jedoch in mindestens einem anderen Schriftsystem funktional lesen und schreiben können, zu einer einzigen, vermeintlich homogenen Gruppe zusammengefasst. Eine weitere Unterscheidung, wie sie beispielsweise bei Analphabeten existiert, findet noch nicht statt. Wie die Resultate implizieren, wäre eine Unterteilung von Zweitschriftlernenden entsprechend ihrer Schulerfahrung jedoch in Betracht zu ziehen, denn diese scheint sich positiv auf das telc-Resultat auszuwirken; Personen mit Schulerfahrung schneiden demnach in einem standardisierten Test besser ab. Bisherige Untersuchungen mit anderen Zielgruppen legen zudem nahe, dass Zweitschriftlernende weiter differenziert werden könnten, und zwar gemäss ihren Schreibkompetenzen in der Erstschriftsprache. Dies lässt sich anhand der vorliegenden Daten nicht bestätigen. Gleiches gilt für das Beherrschen einer oder mehrerer Fremdsprachen.

Zwar konnten im Rahmen der Qualifizierungsarbeit aufschlussreiche Erkenntnisse gewonnen werden, deutlich wird aber vor allem auch, dass breite empirische Evidenz in diesem Forschungsfeld nach wie vor fehlt – und, in Anbetracht erheblicher praktischer Implikationen – dringlich scheint.

9.2 Bildungspolitik

In der Schweiz existieren im Gegensatz zu den deutschsprachigen Nachbarstaaten aktuell keine landesweit geltenden konzeptuellen Grundlagen oder Vorgaben für Zweitschriftlernende oder Analphabeten. Ein entsprechendes Rahmencurriculum für Zweitschriftlerner oder zumindest die Integration der Zielgruppe in ein Alphabetisierungskonzept würde die Grundlage dafür schaffen, dass sie in allen Landesteilen eine bedarfsgerechte Sprachförderung erhält.

Entscheidend ist – in diesem Punkt ist die Evidenz eindeutig – eine klare Trennung zwischen Zweitschriftlernenden und Analphabeten, da beide Zielgruppen unterschiedliche Voraussetzungen für das Sprachenlernen aufweisen. Wie bereits besprochen, gilt es zudem, eine weitere Differenzierung der Zweitschriftlernenden, z. B. anhand ihrer Schulerfahrung, ins Auge zu fassen. Ein eigens für die Schweiz entwickeltes Zweitschriftlernerkonzept sowie die Bereitstellung von Instrumenten und Vorgaben wäre geeignet, die Sprachstandabklärung zu harmonisieren und sie neuesten wissenschaftlichen Erkenntnissen entsprechend auszugestalten. So zeigen die Ergebnisse der vorliegenden Studie, dass das (tatsächlich) erreichte Niveau wesentlich vom Einstufungsniveau abhängt. Während das Ziel A1 innert 300 Lektionen für die besseren Teilnehmenden realistisch ist, scheint es für die niedrigeren Niveaustufen zu ehrgeizig zu sein. Für diese sollten daher alternative Kurskonzepte in Erwägung gezogen werden. Eine korrekte Einstufung ist insbesondere darum notwendig, weil sich ein vorgängig besuchter Deutschkurs als ausgeprägt negativ für den Lernerfolg und die Niveauerreichung erweist, auch wenn in dieser Untersuchung die Hintergründe für diesen Zusammenhang nicht abschliessend geklärt werden konnten.

Wenn ein Zweitschriftlernerkurs, wie er im Pilotprojekt durchgeführt wurde, in der Schweiz zukünftig weiter stattfinden soll, sollten die potenziellen Folgekurse ebenfalls berücksichtigt werden. Kursteilnehmende, die das Niveau A1 erreicht haben, könnten grundsätzlich in reguläre A2-Kurse eintreten. Hier müsste jedoch geprüft werden, ob das hohe Lerntempo und die regelmässige Frequenz aus dem Pilotkurs gewinnbringend übernommen werden können.³⁹ Schwieriger ist es, wenn Personen die A1-Prüfung nicht bestanden haben. Hier müsste im konkreten Fall abgeklärt werden, ob das Nichtbestehen möglicherweise auf mangelnde Lese- und/oder Schreibkompetenzen im lateinischen Schriftsystem zurückzuführen ist. Entsprechend müssten passende A1-Folgekurse oder Lese- und Schreibkurse angeboten werden, wobei Letztere aus Teilnehmerperspektive als demotivierend wahrgenommen werden könnten. Zumindest legt dies der vorgefundene negative Zusammenhang zwischen vorgängig besuchttem Deutschkurs und Lernprogression nahe.

Nicht zuletzt müssten im Rahmenkonzept Qualitätsstandards für Sprachförderangebote erstellt werden, die unter anderem die Ausbildung der Lehrpersonen miteinschliessen und

³⁹ In der durchgeführten Folgestudie wurde unter anderem dieser Frage nachgegangen, Guerrero Calle (im Erscheinen).

entsprechende Erhöhungen in der Bezahlung der Lehrpersonen mit sich ziehen, wie dies bereits vor zehn Jahren gefordert wurde (vgl. Lenz et al. 2009, S. 63).

In jedem Fall ist für die Ausarbeitung eines schweizerischen Rahmenkonzepts für Zweitschriftlernende weitere Forschung nötig. Auch sollte hier die Bildungspolitik Mitverantwortung übernehmen. Ein Interesse an der Thematik und eine entsprechende Finanzierung relevanter Forschungsprojekte würden helfen, komfortablere und nützliche empirische Grundlagen zu schaffen.

9.3 Praxis

Solange kein schweizweites Konzept existiert und es an konkreten Vorgaben fehlt, obliegt es den Kursleitenden und den Sprachförderinstitutionen, den Bedarf an spezifischen Zweitschriftlernerkursen zu erkennen und entsprechende Kurse zu entwickeln und anzubieten. Dazu gehört auch eine korrekte Sprachstandabklärung und Einstufung der Teilnehmenden. Wie bereits mehrfach hervorgehoben, ist die korrekte Einstufung der Lernenden von grundsätzlicher Bedeutung, da sie den möglichen Lernerfolg massgeblich beeinflusst. Wo keine adäquate Sprachstandabklärung angeboten werden kann, sollten sich die Institutionen das entsprechende Instrumentarium und Wissen aneignen oder die Aufgabe anderen Institutionen übertragen, die über entsprechende Instrumente verfügen.

Um Kurse zur Sprachförderung von Migranten zu leiten, wird derzeit keine spezifische Ausbildung verlangt. Dies ungeachtet des Faktums, dass Unterrichten unter Berücksichtigung der aktuellen didaktischen Standards und gemäss dem Bedarf der Kursteilnehmenden nur durch eine entsprechende Aus- bzw. Weiterbildung gewährleistet werden. Während Kursleitende dazu angehalten sind, freiwillig an entsprechenden Bildungsmassnahmen teilzunehmen, könnten Institutionen zur Verbesserung der Situation beitragen, indem sie eine geeignete Vor- bzw. Ausbildung entweder voraussetzen oder diese über ein entsprechendes Entlohnungssystem zumindest attraktiver machen.

Bezüglich der Silbenmethode kann an dieser Stelle keine eindeutige Empfehlung ausgesprochen werden. Als durchgehende Unterrichtsmethode hat sie zwar einen positiven und statistisch signifikanten Einfluss auf die A1-Prüfung, auch wenn dieser relativ gering ist. Für das Hörverstehen bei Teilnehmenden mit wenig Vorwissen scheint sie sich jedoch als hilfreich zu erweisen.

10. Schlusswort

Zweitschriftlernende werden in der deutschsprachigen Schweiz gemeinsam mit Analphabeten in Alphabetisierungskursen unterrichtet. Dieses Vorgehen ist nicht optimal, da Zweitschriftlernende auf deutlich mehr Lernerfahrung zurückgreifen können. Daher stellt sich die Frage, ob eigens für die Zielgruppe geschaffene Kursgefässe effizienter wären.

Aus diesem Grund wurde in der vorliegenden Studie untersucht, ob das im *Zweitschriftlernerkonzept* aus Deutschland beschriebene Kursformat auch für die Schweiz geeignet ist. Ziel dieser Arbeit war es herauszufinden, inwieweit Zweitschriftlernende, wie im deutschen *Zweitschriftlernerkonzept* postuliert, innerhalb von 300 Kurslektionen das Niveau A1 erreichen können und ob die Silbenmethode zur Zielerreichung beitragen kann. Dazu wurden insgesamt 57 Zweitschriftlernende aus fünf verschiedenen Ländern in eigens für sie konzipierten Kursen unterrichtet und hinsichtlich ihres Lernfortschritts untersucht.

Um die genannte Fragestellung abzuklären, wurde eine Längsschnittmessung durchgeführt: Zum einen wurden vor Kursstart sowie nach 100 und nach 200 Kurslektionen forschergenerierte Tests, darunter Elicited Imitation, Kunstwörter, Oral Reading Fluency, Hörverstehen und Diktat, eingesetzt. Zum andern wurde als Abschlussprüfung die standardisierte telc-Prüfung A1 durchgeführt.

Die statistische Auswertung der Daten liefert Hinweise darauf, dass die Erreichung des A1-Niveaus mittels telc zwar für eine gewisse Anzahl ($n = 19$) von Teilnehmenden möglich ist, jedoch nicht für die gesamte Stichprobe. Ob das Zielniveau erreicht wird, ist abhängig von verschiedenen Faktoren, beispielsweise dem Einstufungsniveau, der Erstsprache oder der Anzahl der absolvierten Schuljahre im Heimatland. Anders ausgedrückt weisen die Ergebnisse darauf hin, dass eine differenziertere Unterscheidung von Zweitschriftlernenden, wie sie beispielsweise bei Analphabeten bereits vorgenommen wird, einen bedarfsgerechteren und potenziell effizienteren Unterricht ermöglichen würde.

Die Silbenmethode konnte vor allem für Zweitschriftlernende mit weniger Vorkenntnissen im Deutschen und vor allem im Hörverstehen positive Ergebnisse zeigen. Aufgrund lediglich schwacher positiver Tendenzen und wegen verschiedener Limitationen, darunter beispielsweise die Variable *Institution*, kann die Frage nach dem Nutzen der Silbenmethode jedoch nicht abschliessend beantwortet werden. Da die Effizienz des Einsatzes der Silbenmethode aber mit den Vorkenntnissen im Deutschen in Zusammenhang zu stehen

scheint, können auch diese Resultate als Hinweis darauf betrachtet werden, dass eine adäquate Differenzierung von Zweitschriftlernenden zielführend wäre.

Auch wenn im Rahmen des Forschungsprojekts aufschlussreiche und grundlegende Erkenntnisse gewonnen werden konnten, bleibt eine Reihe von Fragen unbeantwortet. Entsprechend ist zu hoffen, dass sich die empirische und auch die konzeptionelle Forschung in Zukunft vermehrt den Zweitschriftlernenden und Analphabeten in der Schweiz und in anderen Ländern zuwenden wird. Lern-, Lese- und Schreibprozesse der Zielgruppe müssen weiter erforscht und genauer verstanden werden, damit das Kurssystem justiert und die Lernbedingungen den Kursteilnehmenden besser angepasst werden können.

Bibliografie

Abraham, Ellen/Linde, Andrea (2009): Alphabetisierung/Grundbildung als Aufgabengebiet der Erwachsenenbildung. In: Tippelt, Rudolf/Hippel, Aiga von (Hg.): Handbuch Erwachsenenbildung/Weiterbildung. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 889–903.

Acevedo, Jennifer et al. (2015): ESL for Adult Literacy Learners (ALL). Ottawa, Ontario: Centre for Canadian Language Benchmarks.

Aguado, Karin (2000): Empirische Fremdsprachenforschung. Ein Plädoyer für mehr Transparenz. In: Aguado, Karin (Hg.): Zur Methodologie in der empirischen Fremdsprachenforschung. Baltmannsweiler: Schneider Verlag Hohengehren. S. 119–131.

Ahlberg, Aija Katriina (2015): From abugida to alphabet: The interplay between phonological awareness and writing system in transfer literacy – the case of Konso language in South West Ethiopia. In: Humanistis-yhteiskuntatieteellinen tiedekunta. Online unter: <https://www.jyu.fi/hytk/fi/laitokset/solki/opiskelu/opiskelijat/ahlberg-en> [27.01.2020].

Albert, Ruth et al. (2015): Alphabetisierung in der Fremdsprache Deutsch: Lehrmethoden auf dem Prüfstand. Marburg: Tectum Verlag.

Albert, Ruth/Marx, Nicole (2016): Empirisches Arbeiten in Linguistik und Sprachlehrforschung: Anleitung zu quantitativen Studien von der Planungsphase bis zum Forschungsbericht. 3., überarb. und aktualisierte Aufl. Tübingen: Narr Francke Attemptobrooks.

Allyn, Frances A./Burt, Jennifer S. (1998): Pinch my wig or winch my pig: Spelling, spoonerisms and other language skills. In: Reading and Writing Vol. 10, S. 51–74.

ALTE (o. J.): Association of Language Testers in Europe (ALTE) – Q-Mark. Online unter: <https://www.alte.org/Setting-Standards> [27.01.2020].

ALTE (2012): Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER. Frankfurt a. M: o. V.

Amnesty International Schweiz (o. J.): Asyl und Migration. Grundlagen und Begriffe. In: Amnesty International Schweiz. Online unter: <https://www.amnesty.ch/de/themen/asyl-und-migration/zahlen-fakten-und-hintergruende/grundlagen-und-begriffe> [27.01.2020].

Asendorpf, Jens B./Neyer, Franz J. (2012): Psychologie der Persönlichkeit. 5. Aufl. Berlin Heidelberg: Springer-Verlag.

Association of Language Testers in Europe (ALTE (Hg.) (2006): Handreichungen für Testautoren: Modul 1. Modelle der Sprachkompetenz. o. V. Online unter: https://www.testdaf.de/fileadmin/Redakteur/Bilder/Aktuelles/2007/ALTE_Deutsche_HR_Vorwort.pdf [27.01.2020].

Association of Language Testers in Europe (ALTE) (Hg.) (2006): Handreichungen für Testautoren: Modul 4. Aspekte der Bewertung und Auswertung. Online unter: https://www.testdaf.de/fileadmin/Redakteur/Bilder/Aktuelles/2007/ALTE_Deutsche_HR_Vorwort.pdf [27.01.2020].

Bachman, Lyle F. (1990): Fundamental considerations in language testing. Oxford: Oxford Univ. Press.

Backhaus, Alex/Rackwitz, Rüdiger-Philipp (2011a): Lesen & Schreiben: Ein Verfahren zur dialogischen Förderdiagnostik in der Alphabetisierung Jugendlicher und Erwachsener – Handreichung –. Siegen: UniPrint Universität Siegen.

Backhaus, Alex/Rackwitz, Rüdiger-Philipp (2011b): Lesen & Schreiben: Ein Verfahren zur dialogischen Förderdiagnostik in der Alphabetisierung Jugendlicher und Erwachsener – Auswertungshinweise –. Siegen: UniPrint Universität Siegen.

Bader, Dina et al. (2013): Integration im Kanton Bern – Migrationsbevölkerung und Integrationsförderung im Fokus. Neuchâtel: SFM - Forum suisse pour l'étude des migrations et de la population. Online unter: https://www.gef.be.ch/gef/de/index/direktion/organisation/soa/publikationen/migration/grundlagen_konzepte.assetref/dam/documents/GEF/SOA/de/Soziales/Publikationen/Migration/Grundlagen/Integration_im_Kanton_Bern_Bericht_de.pdf [27.01.2020].

BAMF (2015): Konzept für einen bundesweiten Integrationskurs. Nürnberg: Bundesamt für Migration und Flüchtlinge. Überarb. Neuaufl. Online unter: https://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kursraeger/KonzepteLeitfaeden/konz-f-bundesw-integrationskurs.pdf?__blob=publicationFile [27.01.2020].

BAMF (2017a): Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende. Nürnberg: Bundesamt für Migration und Flüchtlinge. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kursraeger/KonzepteLeitfaeden/konzept-zweitschriftlernende.pdf?__blob=publicationFile [27.01.2020].

BAMF (2017b): Die nachstehenden allgemeinen Nebenbestimmungen zur Durchführung der Integrationskurse sind Bestandteil des Zulassungsbescheids. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kursraeger/Traegerrundschreiben/Anlagen/traegerrundschreiben-19_20171228-anlage1.pdf?__blob=publicationFile [27.01.2020].

BAMF (2017c): Merkblatt zum Integrationskurs. Nürnberg: BAMF. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kursraeger/Kursraeger/Merkblaetter/630-009_merkblatt-zum-antrag-auf-zulassung.pdf?__blob=publicationFile [27.01.2020].

BAMF (2018a): Konzept für einen bundesweiten Integrationskurs für Zweitschriftlernende (Zweitschriftlernerkurs). Nürnberg: Bundesamt für Migration und Flüchtlinge. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurst-raeager/KonzepteLeitfaeden/konzept-zweitschriftlernende.pdf?__blob=publicationFile [27.01.2020].

BAMF (2018b): Liste der zugelassenen Lehrwerke in Integrationskursen. Online unter: https://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Lehrkraefte/liste-zugelassener-lehrwerke.pdf?__blob=publicationFile [27.01.2020].

BAMF (2018c): BAMF – Zulassung von Lehrkräften in Integrationskursen. In: BAMF. Online unter: <http://www.bamf.de/DE/Infothek/Lehrkraefte/Zulassung/zulassung.html> [27.01.2020].

BAMF (2018d): Zulassungskriterien für Lehrkräfte in Integrationskursen (§ 15 Abs. 1 und 2 IntV). Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Lehrkraefte/matrix-zulassung-lehrkraefte-integrationskurse.pdf;jsessionid=9612C15C7164CD30CB7B8DF233A9226A.2_cid368?__blob=publicationFilehttp://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Lehrkraefte/matrix-zulassung-lehrkraefte-integrationskurse.pdf;jsessionid=9612C15C7164CD30CB7B8DF233A9226A.2_cid368?__blob=publicationFile [27.01.2020].

BAMF (2019): Zusatzqualifizierungen. In: Bundesamt für Migration und Flüchtlinge. Online unter: http://erwachsenenbildung.at/aktuell/nachrichten_details.php?nid=11231 [27.01.2020].

BAMF (2019a): Richtlinien zur Abrechnung von Integrationskursen vom 20.03.2017. Online unter: <http://www.deutsche-islam-konferenz.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurstraeger/Abrechnung/Richtlinien/richtlinien-abrechnung-stand-20170606.html> [27.01.2020].

BAMF (2019b): Richtlinien zur Abrechnung von Integrationskursen vom 01.01.2019. Online unter: <http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurstraeger/Abrechnung/Richtlinien/richtlinien-abrechnung-stand-20190101.html> [27.01.2020].

BAMF (2019c): Das Bundesamt in Zahlen 2018. Asyl. Nürnberg: BAMF. Online unter: <http://www.bamf.de/DE/Infothek/Statistiken/statistiken-node.html> [27.01.2020].

Barkowski, Hans (2010): Fachlexikon Deutsch als Fremd- und Zweitsprache. Tübingen: Francke.

Bassetti, Benedetta/Vaid, Jyotsna/Cook, Vivian (2012): Interdisciplinary approaches to second language writing systems. In: Writing Systems Research Vol. 4, S. 1–7.

Baten, Kristof (2015): Methodologische Herausforderungen der empirischen Fremdsprachenerwerbsforschung am Beispiel des Kasuserwerbs: Die „Elicited Imitation Task“. In: Zeitschrift für Fremdsprachenforschung Vol. 26, S. 161–181.

- Becker, Thomas (2012): Einführung in die Phonetik und Phonologie des Deutschen. Darmstadt: WBG.
- Becker, Tabea (2013): Schriftspracherwerb in der Zweitsprache: eine qualitative Längsschnittstudie. 2. korr. Aufl. Münster: Schneider Verlag Hohengehren.
- Berthele, Raphael (2018): Alle Thesen sind Hypothesen: Transfer. In: *Babylonia*, S. 63–67.
- Bigelow, Martha/Lovrien Schwarz, Robin (2010): Adult English Language Learners with Limited Literacy. Washington D.C.: o. V.
- Bley-Vroman, Robert/Chaudron, Craig (1994): Elicited Imitation as a measure of second-language competence. In: Tarone, Elaine E./Gass, Susan M./Cohen, Andrew D (Hg.): *Research Methodology in Second-Language Acquisition*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers. S. 245–261.
- Böhme, Gert/Arnold, Barbara (2008): Auditive Verarbeitungs- und Wahrnehmungsstörungen (AVWS) im Kindes- und Erwachsenenalter: Defizite, Diagnostik, Therapiekonzepte, Fallbeschreibungen. 2. Aufl. Bern: Huber.
- Bortz, Jürgen/Döring, Nicola (2006): Forschungsmethoden und Evaluation: für Human- und Sozialwissenschaftler. 4., überarb. Aufl. Heidelberg: Springer.
- Böttinger, Anja (2017): Schritte plus Alpha kompakt. München: Hueber Verlag.
- Bowden, Harriet (2016): Assessing second-language oral proficiency for research: The Spanish Elicited Imitation Task. In: *Studies in Second Language Acquisition* Vol. 38, S. 647–675.
- Brinitzer, Michaela et al. (2013): Fertigkeiten Hören. In: *DaF unterrichten: Basiswissen Didaktik Deutsch als Fremd- und Zweitsprache*. Stuttgart: Klett Sprachen. S. 24–34.
- Brooks, Greg et al. (2001): *Progress in Adult Literacy: Do Learners Learn?* London: Basic Skills Agency.
- Buck, Gary (2001): *Assessing listening*. Cambridge: Cambridge University Press.
- Buetler, Karin Andrea et al. (2014): Language context modulates reading route: an electrical neuroimaging study. In: *Frontiers in Human Neuroscience* Vol. 8, S. 1–16.
- Bühner, Markus (2011): Einführung in die Test- und Fragebogenkonstruktion. 3., aktualisierte und erw. Aufl. München: Pearson Studium.
- Bühner, Markus/Bernd, Marcus (2009): *Grundlagen der Testkonstruktion*. Hagen: FernUniversität in Hagen.
- Bulut, Neclet et al. (2010): Vom Buchstaben zum Text – Ein adaptiver Test zur Messung von Schriftsprachstand und -entwicklung von erwachsenen Lernern. In: Roll, Heike/Schramm, Karen (Hg.): *Alphabetisierung in der Zweitsprache Deutsch*. Oldenburg: Redaktion OBST. S. 127–142.

Bundesministerium Inneres (2018): Asylstatistik 2018. Wien: Bundesministerium Inneres. Online unter: <https://www.bmi.gv.at/301/Statistiken/> [27.01.2020].

Burt, Miriam/Peyton, Joy Kreeft (2003): Reading and Adult English Language Learners: The Role of the First Language. Washington D.C.: Center for Applied Linguistics.

Caroni, Martina et al. (Hg.) (2014): Migrationsrecht. 3., stark überarb. Aufl. Bern: Stämpfli.

Castro-Caldas, Alexandre et al. (1998): The illiterate brain: Learning to read and write during childhood influences the functional organization of the adult brain. In: Brain Bd. 121, S. 1053–1063.

Ciuffo, Massimo et al. (2017): How fast can we read in the mind? Developmental trajectories of silent reading fluency. In: Reading and Writing Vol. 30, S. 1667–1686.

Cohen, Jacob (1988): Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (N.J.): Lawrence Erlbaum.

Cook, Vivian (2005): Written Language and Foreign Language Teaching. In: Cook, Vivian/Basseti, Benedetta (Hg.): Second Language Writing Systems. Clevedon; Buffalo: Multilingual Matters. S. 424–441.

Cossu, Giuseppe/Gugliotta, Maria/Marshall, John C. (1995): Acquisition of reading and written spelling in a transparent orthography: Two non parallel processes? In: Reading and Writing Vol. 7, S. 9–22.

Dahmen, Silvia/Weth, Constanze (2018): Phonetik, Phonologie und Schrift. Paderborn: Schöningh.

Darsow, Annkathrin/Felbrich, Anja (2014): Besondere Forschungsansätze: Experiment und Quasi-Experiment. In: Settinieri, Julia et al. (Hg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: UTB. S. 229–241.

Davidson, Rosalind Kastle/Strucker, John (2002): Patterns of word recognition errors among adult basic education native and non-native speakers of English. In: Scientific Studies of Reading Vol. 6, S. 299–316.

Der Bundesrat (2018): Abkommen über die Rechtsstellung der Flüchtlinge. In: Schweizerische Eidgenossenschaft. Online unter: <https://www.admin.ch/opc/de/classified-compilation/19510156/> [27.01.2020].

Derichs-Kunstmann, Karin/Chiersmann, Christiane/Tippelt, Rudolf (Hg.) (1993): Die Fremde – das Fremde – der Fremde: Dokumentation der Jahrestagung 1992 der Kommission Erwachsenenbildung der Deutschen Gesellschaft für Erziehungswissenschaft. Frankfurt a. M: Pädagogische Arbeitsstelle des DVV.

Döbert, Marion/Hubertus, Peter (2000): Ihr Kreuz ist die Schrift: Analphabetismus und Alphabetisierung in Deutschland. Münster: Bundesverband Alphabetisierung.

Drackert, Anastasia (2015): Elicited Imitation. In: Validating Language Proficiency Assessments in Second Language Acquisition Research: Applying an Argument-Based Approach. New edition. Frankfurt am Main; New York: Peter Lang GmbH, Internationaler Verlag der Wissenschaften. S. 49–67.

Dürscheid, Christa (2016): Einführung in die Schriftlinguistik. 5., aktualisierte und korr. Aufl. Göttingen: UTB.

Edmondson, Willis J/House, Juliane (2011): Einführung in die Sprachlehrforschung. überarb. Aufl. Tübingen und Basel: A. Francke Verlag.

Egloff, Birte et al. (2011): Funktionaler Analphabetismus im Erwachsenenalter: eine Definition. In: Projektträger im DLR e.V. (Hg.): Zielgruppen in Alphabetisierung und Grundbildung Erwachsener: Bestimmung, Verortung, Ansprache. Bielefeld: W. Bertelsmann. S. 11–31.

Erlam, Rosemary (2006): Elicited Imitation as a Measure of L2 Implicit Knowledge: An Empirical Validation Study. In: Applied Linguistics Vol. 27, S. 464–491.

Esser, Hartmut (2006a): Sprache und Integration. Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten. Frankfurt/New York: Campus Verlag.

Esser, Hartmut (2006b): Migration, Sprache und Integration. Berlin: Arbeitsstelle Interkulturelle Konflikte und gesellschaftliche Integration (AKI), Wissenschaftszentrum Berlin für Sozialforschung (WZB). Online unter: <http://www.ssoar.info/ssoar/bitstream/handle/document/11349/ssoar-2006-esser-migration.pdf?sequence=1> [27.01.2020].

Feick, Diana/Schramm, Karen (2016): Alphabetisierung mit Migrantinnen und Migranten. In: Löffler, Cordula/Korfkamp, Jens (Hg.): Handbuch zur Alphabetisierung und Grundbildung Erwachsener. Münster, New York: Waxmann. S. 214–225.

Feldmeier, Alexis (2005a): Die kontrastive Alphabetisierung als Alternativkonzept zur zweisprachigen Alphabetisierung und zur Alphabetisierung in der Zweitsprache Deutsch am Beispiel der Sprachen Kurdisch und Türkisch. In: Deutsch als Zweitsprache, S. 42–50.

Feldmeier, Alexis (2005b): Die Bedeutung der Deutschkenntnisse für die Alphabetisierung multinationaler Lernergruppen. Am Beispiel der Übungsform „Diktat“. In: Deutsch als Zweitsprache, S. 8–15.

Feldmeier, Alexis (2007): Vorläufiges Konzept für einen bundesweiten Integrationskurs mit Alphabetisierung. Nürnberg: BAMF. Online unter: https://www.alphabetisierung.de/fileadmin/files/Bilder/News/konzept-fuer-einen-bundesweiten-integrationskurs-mit-alphabetisierung_IP.pdf [27.01.2020].

Feldmeier, Alexis (2008): Möglichkeiten und Grenzen des „Vorläufigen Konzepts für einen bundesweiten Integrationskurs mit Alphabetisierung“. In: Knabe, Ferdinande/Bundesverband Alphabetisierung und Grundbildung e.V. (Hg.): Innovative Forschung – innovative Praxis in der Alphabetisierung und Grundbildung. Münster: Waxmann Verlag. S. 176–192.

Feldmeier, Alexis (2010a): Von A bis Z – Praxishandbuch Alphabetisierung Deutsch als Zweitsprache für Erwachsene. Stuttgart: Ernst Klett Sprachen.

Feldmeier, Alexis (2010b): Alphabetisierung von Erwachsenen nicht deutscher Muttersprache. Leseprozesse und Anwendung von Strategien beim Erlesen isoliert dargestellter Wörter unter besonderer Berücksichtigung der farblichen und typographischen Markierung von Buchstabengruppen. Bielefeld: Universität Bielefeld. Online unter: <https://pub.uni-bielefeld.de/publication/2301752> [27.01.2020].

Feldmeier, Alexis (2010c): Von A bis Z – Praxishandbuch Alphabetisierung: Deutsch als Zweitsprache für Erwachsene. Stuttgart: Ernst Klett Sprachen.

Feldmeier, Alexis (2010d): Von A bis Z – Alphabetisierungskurs A1: Deutsch als Zweitsprache für Erwachsene. Stuttgart: Ernst Klett Sprachen.

Feldmeier, Alexis (2015): Konzept für einen bundesweiten Alphabetisierungskurs. Nürnberg: Bundesamt für Migration und Flüchtlinge. Überarb. Neuaufl. Online unter: <http://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Integrationskurse/Kurst-raeager/KonzepteLeitfaeden/konz-f-bundesw-ik-mit-alphabet.pdf> [27.01.2020].

Fellmer, Tim-Thilo/Feldmeier, Alexis (2012): Von A bis Z – Leseheft für den Alphabetisierungskurs. Lesestufe 1. Stuttgart: Ernst Klett Sprachen.

Florida Department of Education (2006): Instructions for Administering the Native Language Literacy Screening Device for placement in Literacy for Adult ESOL Learners course. Florida: Florida Department of Education. Online unter: <https://cehdvision2020.umn.edu/wp-content/uploads/2013/08/vocscr.pdf> [27.01.2020].

Florida Department of Education (2014): Native Language Literacy Screening Manual. Florida: Florida Department of Education. Online unter: <http://www.fldoe.org/core/fileparse.php/7522/urlt/0061376-native.pdf> [27.01.2020].

Flowerdew, John/Miller, Lindsay (2014): Assessing Listening. In: Coombe, Christine A. et al. (Hg.): The Cambridge guide to second language assessment. New York: Cambridge University Press. S. 225–233.

Frith, Uta (1985): Beneath the Surface of Developmental Dyslexia. In: Patterson, Karalyn E./Marshall, John C./Coltheart, Max (Hg.): Surface dyslexia: neuropsychological and cognitive studies of phonological reading. London, Hillsdale: L. Erlbaum Associates Publ. S. 301–330.

Fritz, Thomas et al. (2006): Rahmencurriculum Deutsch als Zweitsprache & Alphabetisierung. Wien: Lehrstuhl Deutsch als Fremdsprache an der Universität Wien / Institut für Weiterbildung / Verband Wiener Volksbildung; AlfaZentrum für MigrantInnen der Volkshochschule Ottakring. Online unter: <https://www.wien.gv.at/menschen/integration/pdf/rahmen-curriculum.pdf> [27.01.2020].

Fuchs, Lynn S. et al. (2001): Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. In: Scientific Studies of Reading Vol. 5, S. 239–256.

Gagliano, Antonella et al. (2015): Silent reading fluency: Implications for the assessment of adults with developmental dyslexia. In: Journal of Clinical and Experimental Neuropsychology Vol. 37, S. 972–980.

Gamper, Hans et al. (2012): Normen für den Mottier-Test bei 4- bis 12-jährigen Kindern. Bern: Praxisforschung der Erziehungsdirektion des Kantons Bern.

Gathercole, Susan E. et al. (1994): The children's test of nonword repetition: A test of phonological working memory. In: Memory Vol. 2, S. 103–127.

Gathercole, Susan E. (2006): Nonword repetition and word learning: The nature of the relationship. In: Applied Psycholinguistics Vol. 27, S. 513–543.

Gehring, Wolfgang (2018): Fremdsprache Deutsch unterrichten: kompetenzorientierte Methoden für DaF und DaZ. Bad Heilbrunn: Verlag Julius Klinkhardt.

Gesundheits- und Fürsorgedirektion des Kantons Bern (2017): Kantonales Integrationsprogramm 2018–2021 Aktionsplan. Online unter: https://www.gef.be.ch/gef/de/index/migration_integration/migration_integration.assetref/dam/documents/GEF/SOA/de/Soziales/Publikationen/Migration/Grundlagen/Kantonales_Integrationsprogramm_KIP_2018_2021_de.pdf [27.01.2020].

Gesundheits- und Fürsorgedirektion GEF und Polizei und Militärdirektion POM (2017): Detailkonzeption Neustrukturierung Asyl- und Flüchtlingsbereich im Kanton Bern. Detailkonzeption zur Umsetzung der Gesamtstrategie für den Asyl- und Flüchtlingsbereich im Kanton Bern gemäss RRB 907/2016 vom 17. August 2016 unter Berücksichtigung der Planungserklärungen des Grossen Rates vom 18. November 2016. Online unter: https://www.asyl.sites.be.ch/asyl_sites/de/index/navi/index/na-be.assetref/dam/documents/POM/MIP/de/MIDI/Internetportal%20Asyl/Detailkonzeption_NA-BE_d.pdf [27.01.2020].

Gillespie, Marilyn K. (2001): Research on Writing: Implications for Adult Literacy Education. In: Comings, John/Garner, Barbara/Smith, Christine (Hg.): Review of Adult Learning and Literacy. San Francisco: Jossey-Bass. S. 63–110.

Goethe Institut (2009): Mehrsprachigkeit und Quersprachigkeit – Kreativpotenzial von Autorinnen und Autoren mit nichtdeutscher Muttersprache. In: Goethe-Institut. Online unter: <http://www.goethe.de/ges/spa/prj/sog/muk/de4714712.htm> [27.01.2020].

Grabe, William/Zhang, Cui (2016): Reading-writing relationships in first and second language academic literacy development. In: Language Teaching Vol. 49, S. 339–355.

Graham, C. Ray et al. (2008): Elicited imitation as an oral proficiency measure with ASR scoring. In: Proceedings of LREC, S. 1604–1610.

Graham, C. Ray/McGhee, Jeremiah/Millard, Ben (2010): The Role of Lexical Choice in Elicited Imitation Item Difficulty. In: Prior, Matthew T./Watanabe, Yukiko/Lee, Sang-Ki (Hg.): Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices. Sommerville, MA: Cascadia Proceedings Project. S. 57–72.

Green, Anthony (2014): Exploring language assessment and testing. Language in action. Abingdon: Routledge.

Grotjahn, Rüdiger (2001): Leistungsmessung und Leistungsbeurteilung. Kapitel 5 Testen der Fertigkeit Hörverstehen. In: Hellenic Open University Patras (Hg.): Leistungsmessung und Leistungsbeurteilung. Patras: o. V. Postgraduiertenstudium in Deutsch als Fremdsprache.

Grotjahn, Rüdiger (2010): Beurteilen und Evaluieren. In: Hallet, Wolfgang/Königs, Frank G. (Hg.): Handbuch Fremdsprachendidaktik. Seelze-Velber: Kallmeyer in Verbindung mit Klett. S. 211–215.

Grum, Urška/Legutke, Michael K. (2016): Sampling. In: Caspari, Daniela et al. (Hg.): Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch. Tübingen: Narr Francke Attempto Verlag GmbH. S. 78–89.

Guerrero Calle, Santi (im Erscheinen): Assessing the impact of the syllabary approach on German literacy gains for Tigrinya non-Roman alphabet literate learners. In: LESLLA. Palermo: Palermo Academic Press (UniPa Press).

Guerrero Calle, Santi/Studer, Thomas/Lenz, Peter (im Erscheinen): Literaturbericht: Alphabetisierung von erwachsenen Migrant/innen. Fribourg: Wissenschaftliches Kompetenzzentrum für Mehrsprachigkeit.

Günther, Klaus B. (1986): Ein Stufenmodell der Entwicklung kindlicher Lese- und Schreibstrategien. In: Brügelmann, Hans (Hg.): ABC und Schriftsprache: Rätsel für Kinder, Lehrer und Forscher. Konstanz: Faude.

Gupta, Prahlad (2003): Examining the Relationship between word Learning, Nonword Repetition, and Immediate Serial Recall in Adults. In: Quarterly Journal of Experimental Psychology Vol. 56, S. 1213–1236.

Hammann, Andrea et al. (2013): Rahmenbedingungen und Grundlagen der Alphabetisierungsarbeit. In: Feick, Diana/Pietzuch, Anja/Schramm, Karen (Hg.): Alphabetisierung für Erwachsene. München: Klett-Langenscheidt. S. 19–33.

Harsch, Claudia (2016): Testen. In: Caspari, Daniela et al. (Hg.): Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch. Tübingen: Narr Francke Attempto Verlag GmbH. S. 205–217.

Hasbrouck, Jan/Tindal, Gerald A. (2006): Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. In: The Reading Teacher Vol. 59, S. 636–644.

Hasselhorn, Marcus/Körner, Katja (1997): Nachsprechen von Kunstwörtern: Zum Zusammenhang zwischen Arbeitsgedächtnis und syntaktischen Sprachleistungen bei Sechs- und Achtjährigen. In: Zeitschrift für Entwicklungspsychologie u. Pädagogische Psychologie Vol. 29, S. 212–224.

Hasselhorn, Marcus/Seidler-Brandler, Ulrich/Körner, Katja (2000): Ist das „Nachsprechen von Kunstwörtern“ für die Entwicklungsdiagnostik des phonologischen Arbeitsgedächtnisses geeignet? In: Hasselhorn, Marcus/Schneider, Wolfgang/Marx, Harald (Hg.): Diagnostik von Lese- Rechtschreibschwierigkeiten. Göttingen: Hogrefe, Verl. für Psychologie. S. 119–133.

Hattie, John (2013): International guide to student achievement. New York: Routledge.

Heyn, Anne (2013): Sprachen lernen mit Methode: Der Rückgriff der Muttersprache im Sprachunterricht. Marburg: Tectum Verlag.

Hubertus, Peter (1995): Wo steht die Alphabetisierungsarbeit heute? In: Brüggemann, Hans/Balhorn, Heiko/Füssenich, Iris (Hg.): Am Rande der Schrift. Zwischen Sprachenvielfalt und Analphabetismus. Lengwil am Bodensee: Libelle. S. 250–262.

Hubertus, Peter/Nickel, Sven (2003): Sprachunterricht in der Erwachsenenbildung: Alphabetisierung von Erwachsenen. In: Bredel, Ursula et al. (Hg.): Didaktik der deutschen Sprache: ein Handbuch. Paderborn, München, Wien, Zürich: Ferdinand Schöningh. S. 719–728.

Hufeisen, Britta/Riemer, Claudia (2010): Spracherwerb und Sprachenlernen. In: Krumm, Hans-Jürgen et al. (Hg.): Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch. Berlin: Walter de Gruyter. Bd. 1, S. 738–753.

Jones, Randall L./Tschirner, Erwin P. (2006): A Frequency Dictionary of German: Core Vocabulary for Learners. London & New York: Routledge.

Jude, Nina/Klieme, Eckhard (2006): Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik. In: Beck, Bärbel/Klieme, Eckhard (Hg.): Sprachliche Kompetenzen: Konzepte und Messung: DESI-Studie (Deutsch Englisch Schülerleistungen-International). Weinheim, Basel: Beltz Verlag. S. 9–22.

Kazazoglu, Semin (2013): Dictation as a Language Learning Tool. In: Procedia - Social and Behavioral Sciences Vol. 70, S. 1338–1346.

Klicpera, Christian/Gasteiger-Klicpera, Barbara/Schabmann, Alfred (1993): Lesen und Schreiben: Entwicklung und Schwierigkeiten: die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit. Bern: Huber.

Klicpera, Christian/Schabmann, Alfred/Gasteiger-Klicpera, Barbara (2013): Legasthenie – LRS: Modelle, Diagnose, Therapie und Förderung. 4., aktualisierte Aufl. München, Basel: E. Reinhardt.

König, Katharina (2016): Erstsprache – Herkunftssprache – Muttersprache. Sprachbiographische Zugriffe von Deutsch-TürkInnen auf den Ausdruck Muttersprache. In: Raml, Monika (Hg.): Wanderer zwischen den Welten: deutsch-türkische SprachBiographien. Würzburg: Königshausen & Neumann. S. 269–294.

- Kuhberg, Heinz (2001): Zweitspracherwerb als präterminierte Entwicklung I: der behavioristische Ansatz. In: Helbig, Gerhard et al. (Hg.): Deutsch als Fremdsprache: Ein internationales Handbuch. New York: Walter de Gruyter. S. 654–663.
- Küppers, Almut (2006): Diktate, Diktieren und diktiert Bearbeiten. In: Jung, Udo O. H. (Hg.): Praktische Handreichung für Fremdsprachenlehrer. 4., vollst. neu bearb. Aufl. Frankfurt a. M: Peter Lang. S. 90–94.
- Kurvers, Jeanne (2015): Emerging literacy in adult second-language learners: A synthesis of research findings in the Netherlands. In: Writing Systems Research Vol. 7, S. 58–78.
- Landgraf, Steffen et al. (2012): Impact of phonological processing skills on written language acquisition in illiterate adults. In: Development Cognitive Neuroscience Vol. 2, S. 129–138.
- Landis, J. Richard/Koch, Gary G. (1977): The Measurement of Observer Agreement for Categorical Data. In: Biometrics Vol. 33, S. 159–174.
- Leki, Ilona/Cumming, Alister/Silva, Tony (Hg.) (2008): A Synthesis of Research on Second Language Writing in English. New York, London: Routledge.
- Lems, Kristin (2003): Adult ESL Oral Reading Fluency and Silent Reading Comprehension. Chicago, Illinois: National Louis University. Online unter: <https://digitalcommons.nlu.edu/cgi/viewcontent.cgi?article=1136&context=diss> [27.01.2020].
- Lenhard, Wolfgang/Lenhard, Alexandra (2017): Berechnung des Lesbarkeitsindex (LIX) nach Björnson. Online unter: <https://www.psychometrica.de/lix.html> [27.01.2020].
- Lenz, Peter/Andrey, Stéphanie/Lindt-Bangerter, Bernhard (2009): Rahmencurriculum für die sprachliche Förderung von Migrantinnen und Migranten. Bern: Bundesamt für Migration BFM. Online unter: <https://www.sem.admin.ch/dam/data/sem/integration/berichte/sprache/rahmencurriculum-d.pdf> [27.01.2020].
- LESLLA (o. J.): Literacy Education and Second Language Learning for Adults. In: Literacy Education and Second Language Learning for Adults. Online unter: <https://www.leslla.org> [27.01.2020].
- Lienert, Gustav Adolf/Raatz, Ulrich (1998): Testaufbau und Testanalyse. 6. Aufl., Studienausg. Weinheim: Beltz, Psychologie Verlags Union.
- Linde, Andrea (2007): Alphabetisierung, Grundbildung oder Literalität? In: Grotlüschen, Anke/Linde, Andrea (Hg.): Literalität, Grundbildung oder Lesekompetenz? Münster: Waxmann. S. 90–99.
- Linde, Andrea (2008): Literalität und Lernen. Eine Studie über das Lesen- und Schreibenlernen im Erwachsenenalter. Münster: Waxmann.
- Löffler, Cordula/Korfkamp, Jens (2016): Einführung. In: Löffler, Cordula/Korfkamp, Jens (Hg.): Handbuch zur Alphabetisierung und Grundbildung Erwachsener. Münster, New York: Waxmann. S. 9–10.

Martin, Katherine I. (2015): L1 impacts on L2 component reading skills, word skills, and overall reading achievement. Pittsburgh: University of Pittsburgh. Online unter: <https://search.proquest.com/docview/1717095010?pq-origsite=gscholar> [27.01.2020].

Marx, Nicole (2012): Experimentelle Fremdsprachenforschung. In: Doff, Sabine (Hg.): Fremdsprachenunterricht empirisch erforschen. Grundlagen – Methoden – Anwendung. Tübingen: Narr Verlag. S. 51–70.

Maurer, Ernst (2017): Deutsch in der Schweiz A1. Ein Sprachkurs für Erwachsene. Kursbuch. 6. unveränderter Nachdruck. Zug: Klett und Balmer.

Mayer, Andreas (Hg.) (2016): Lese-Rechtschreibstörungen (LRS). München, Basel: E. Reinhardt.

Mayer, Andreas (2018): Gezielte Förderung bei Lese- und Rechtschreibstörungen. 3., überarb. Aufl. München: Ernst Reinhardt Verlag.

McDade, Hiram L./Simpson, Martha A./Lamb, Donna Elmer (1982): The Use of Elicited Imitation as a Measure of Expressive Grammar. In: Journal of Speech and Hearing Disorders Vol. 47, S. 19–24.

Mellard, Daryl F./Anthony, Jason L./Woods, Kari L. (2012): Understanding oral reading fluency among adults with low literacy: dominance analysis of contributing component skills. In: Reading and Writing Vol. 25, S. 1345–1364.

Mellard, Daryl/Woods, Kari/Fall, Emily (2011): Assessment and Instruction of Oral Reading Fluency among Adults with Low Literacy. In: Adult Basic Education and Literacy Journal Vol. 5, S. 1–14.

Müller, Astrid (2010): Rechtschreiben lernen: die Schriftstruktur entdecken – Grundlagen und Übungsvorschläge. Seelze: Kallmeyer.

National Academics of of Sciences, Engineering and Medicine (2020): A Principled Approach to Language Assessment: Considerations for the U.S. Foreign Service Institute. Online unter: <https://www.nap.edu/download/25748> [10.06.2020].

New York State Education Department (1999): Native Language Literacy Screening Device. Glenmont: The University of the State of New York. Online unter: http://www.ohioable.org/files/PDEventMaterials/2013%20Spring%20PD%20Day/Native_Language_Literacy_Screening_Device/intro1.pdf [27.01.2020].

Nickel, Sven (1998): Zugriffe funktionaler Analphabeten auf Schrift. Eine Untersuchung von Schreibstrategien mit der „Hamburger Schreib-Probe“. In: Alfa-Forum Vol. 38, S. 20–24.

Nickel, Sven (2006): Orthographieerwerb und die Entwicklung von Sprachbewusstheit: zu Genese und Funktion von orthographischen Bewusstseinsprozessen im frühen Rechtschreiberwerb in unterschiedlichen Lernkontexten. Norderstedt: Books on Demand GmbH.

Nickel, Sven (2011): Textschwierigkeit objektivieren: Der Lesbarkeitsindex LIX. Wie schwierig sind Lesetexte in der Alphabetisierung? In: Alfa-Forum Vol. 76, S. 30–32.

Niebisch, Daniela et al. (2017): Schritte plus Neu: Deutsch als Zweitsprache. München: Hueber Verlag.

Nold, Günter/Henning, Rossa (2007): Hörverstehen. In: Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim, Basel: Beltz Verlag. S. 178–196.

Nold, Günter/Willenberg, Heiner (2007): Lesefähigkeit. In: Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim, Basel: Beltz Verlag. S. 23–41.

Odlin, Terence (1989): Language transfer: cross-linguistic influence in language learning. Cambridge: Cambridge University Press.

OECD (Hg.) (1998): Grundqualifikationen, Wirtschaft und Gesellschaft: Ergebnisse der ersten internationalen Untersuchung von Grundqualifikationen Erwachsener. Paris: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD).

Olariu, Antonia (2007): Individuelle Mehrsprachigkeit und begriffliche Gegenüberstellung von: Erstsprache, Muttersprache, Zweitsprache und Fremdsprache. In: Philologica Jassyensia Vol. 2, S. 301–306.

Ortega, Lourdes et al. (2002): An investigation of elicited imitation tasks in crosslinguistic SLA research. Toronto. Online unter: <https://www.iris-database.org/iris/app/home/detail?id=york%3a852671&ref=search> [27.01.2020].

Panagiotopoulou, Argyro (2001): Analphabetismus in literalen Gesellschaften am Beispiel Deutschlands und Griechenlands. Frankfurt a.M: Peter Lang GmbH.

Perfetti, Charles A./Dunlap, Susan (2008): Learning to read: General principles and writing system variations. In: Koda, Keiko/Zehler, Annette M. (Hg.): Learning to Read Across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development. New York, NY: Routledge. S. 13–38.

Perlmann-Balme, Michaela/Dengler, Stefanie (2007): Einstufungssystem für die Integrationskurse in Deutschland. Handreichung für Einstufende. Goethe Institut.

Petermann, Franz/Daseking, Monika (2018): ZLT-II. Zürcher Lesetest – II. Weiterentwicklung des Zürcher Lesetests (ZLT) von Maria Linder und Hans Grisseemann. Manual. 3. überarb. Aufl. mit erweiterten Normen. Bern: Hogrefe.

Petscher, Yaacov/Kim, Young-Suk (2011): The utility and accuracy of oral reading fluency score types in predicting reading comprehension. In: Journal of School Psychology Vol. 49, S. 107–129.

Porsch, Raphaela/Grotjahn, Rüdiger/Tesch, Bernd (2010): Hörverstehen und Hör-Sehverstehen in der Fremdsprache – unterschiedliche Konstrukte? In: Zeitschrift für Fremdsprachenforschung Vol. 21, S. 143–189.

- Pretorius, Elizabeth J./Spaull, Nic (2016): Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa. In: *Reading and Writing* Vol. 29, S. 1449–1471.
- Rackwitz, Rüdiger-Philipp (2016): Pseudowörter als Diagnosehilfe in der Alphabetisierung Jugendlicher und Erwachsener: Entwicklung und Erprobung eines Verfahrens zur Lernstandserhebung und Lernverlaufsbeobachtung. Schwäbisch Gmünd: Pädagogische Hochschule Schwäbisch Gmünd.
- Rahimi, Mohammad (2008): Using Dictation to Improve Language Proficiency. In: *The Asian EFL Journal* Vol. 10, S. 33–47.
- Röder, Antje et al. (2018): Methodische Herausforderungen quantitativer Befragungen von Geflüchteten am Beispiel einer Vorstudie in Sachsen. In: *Zeitschrift für Flüchtlingsforschung, Z'Flucht* Vol. 2, S. 313–329.
- Rokitzki, Christiane (2016): Alphabetisierung von erwachsenen Migranten nach Montessori. Ein methodischer Ansatz für die Fremd- und Zweitsprache Deutsch. Marburg: Tectum Verlag.
- Rokitzki, Christiane/Nestler, Doreen/Sokolowsky, Celia (2013): Lernabenteuer Schriftsystem. In: Feick, Diana/Pietzuch, Anja/Schramm, Karen (Hg.): *Alphabetisierung für Erwachsene*. München: Klett-Langenscheidt. S. 91–108.
- Rosebrock, Cornelia et al. (2017): Leseflüssigkeit fördern: Lautleseverfahren für die Primar- und Sekundarstufe. 5. Aufl. Seelze: Klett Kallmeyer.
- Rossa, Henning (2012): Mentale Prozesse beim Hörverstehen in der Fremdsprache: eine Studie zur Validität der Messung sprachlicher Kompetenzen. Frankfurt am Main: P. Lang.
- Rother, Nina (2010): Das Integrationspanel: Ergebnisse einer Befragung von Teilnehmenden zu Beginn ihres Alphabetisierungskurses. Nürnberg: Bundesamt für Migration und Flüchtlinge. Working Paper 29 der Forschungsgruppe des Bundesamtes. Online unter: <http://nbn-resolving.de/urn:nbn:de:0168-ss0ar-259111> [27.01.2020].
- Scheible, Jana A./Rother, Nina (2017): Schnell und erfolgreich Deutsch lernen – wie geht das? Erkenntnisse zu den Determinanten des Zweitspracherwerbs unter besonderer Berücksichtigung von Geflüchteten. Nürnberg: Bundesamt für Migration und Flüchtlinge. Working Paper 72 der Forschungsgruppe des Bundesamtes. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/WorkingPapers/wp72-erfolgreich-deutsch-lernen-wie.pdf?__blob=publicationFile [27.01.2020].
- Scheibner-Herzig, Gudrun/Sauerbrey, Heike/Kokoschka, Siegfried (1991): Repetition – a means to predict foreign language oral proficiency. In: *International Review of Applied Linguistics in Language Teaching* Vol. 29, S. 229–257.
- Scheithauer, Christiane (2010): Diagnose phonologischer Bewusstheit in der Zweitsprache – Erprobung im Integrationskurs mit Alphabetisierung. In: Bolte, Henning/Roll, Heike/Schramm, Karen (Hg.): *Alphabetisierung in der Zweitsprache Deutsch*. Duisburg: Redaktion Obst. S. 75–90.

Scheithauer, Christiane/Waggershauser, Elena/Becker, Tabea (2013): Phonologische Bewusstheit. In: Feick, Diana/Pietzuch, Anja/Schramm, Karen (Hg.): Alphabetisierung für Erwachsene. München: Klett-Langenscheidt. S. 64–78.

Scherf, Daniel/Rieckmann, Carola (2016): Lautleseprotokolle. Diagnose- und Feedbackinstrument in der Erwachsenen-Alphabetisierung. In: Alfa-Forum Vol. 90, S. 32–37.

Schnitzler, Carola D. (2008): Phonologische Bewusstheit und Schriftspracherwerb. Stuttgart: Thieme.

Schoonen, Rob et al. (2003): First Language and Second Language Writing: The Role of Linguistic Knowledge, Speed of Processing, and Metacognitive Knowledge. In: Language Learning: A Journal of Research in Language Studies Vol. 53, S. 165–202.

Schründer-Lenzen, Agi (2013): Schriftspracherwerb. 4., völlig überarb. Aufl. Wiesbaden: Imprint: Springer VS, Imprint: Springer VS.

Schuller, Karin/Lochner, Susanne/Rother, Nina (2011): Das Integrationspanel – Ergebnisse einer Längsschnittstudie zur Wirksamkeit und Nachhaltigkeit von Integrationskursen. Nürnberg: Bundesamt für Migration und Flüchtlinge. Online unter: http://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/Forschungsberichte/fb11-integrationspanel.pdf?__blob=publicationFile [27.01.2020].

Schuller, Karin/Lochner, Susanne/Rother, Nina (2012): Das Integrationspanel: Entwicklung der Deutschkenntnisse und Fortschritte der Integration bei Teilnehmenden an Alphabetisierungskursen. Nürnberg: Bundesamt für Migration und Flüchtlinge. Online unter: https://www.bamf.de/SharedDocs/Anlagen/DE/Publikationen/WorkingPapers/wp42-integrationspanel.pdf?__blob=publicationFile [27.01.2020].

Seitz-Stein, Katja et al. (2012): Diagnostik der Funktionstüchtigkeit des Arbeitsgedächtnisses bei Kindern zwischen 5 und 12 Jahren: Die Arbeitsgedächtnistestbatterie AGTB 5-12. In: Hasselhorn, Marcus/Zoelch, Christof (Hg.): Funktionsdiagnostik des Arbeitsgedächtnisses. Göttingen: Hogrefe.

Sendlmeier, Walter F./Oertel, Alexandra (2015): Rechtschreibdidaktiken im ersten Schuljahr: eine psychologische und sprachwissenschaftliche Einordnung und Bewertung. Berlin: Logos.

Siekman, Katja/Thomé, Günther (2018): Der orthographische Fehler: Grundzüge der orthographischen Fehlerforschung und aktuelle Entwicklung. 2., aktualisierte Aufl. [Oldenburg]: Isb, Institut für sprachliche Bildung Verlag.

Solmecke, Gert (2010): Vermittlung der Hörfertigkeit. In: Krumm, Hans-Jürgen et al. (Hg.): Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch. Berlin: Walter de Gruyter. Bd. 1, S. 969–975.

Spruck Wrigley, Heide (2008): Adult ESL and Literacy: Issues and Options. Montreal: The Centre for Literacy. Online unter: <https://eric.ed.gov/?id=ED547395> [27.01.2020].

Staatssekretariat für Migration SEM (2011): Aufenthaltsbewilligungen für EU/EFTA-Angehörige. Online unter:
https://www.sem.admin.ch/sem/de/home/themen/aufenthalt/eu_efta.html [27.01.2020].

Staatssekretariat für Migration SEM (2019a): Asylstatistik 2. Quartal 2019. Bern: Staatssekretariat für Migration SEM. Online unter:
<https://www.sem.admin.ch/sem/de/home/aktuell/news/2019/2019-08-15.html> [27.01.2020].

Staatssekretariat für Migration SEM (Hg.) (2019b): Kurzinformationen. Bundes Publikationen. Online unter:
<https://www.sem.admin.ch/dam/data/sem/publiservice/publikationen/info-flue-va/info-flue-va-de.pdf> [27.01.2020].

Stetter, Christian (2005): System und Performanz: symboltheoretische Grundlagen von Medientheorie und Sprachwissenschaft. Weilerswist: Velbrück Wissenschaft.

Sturm, Afra (2011): Leseflüssigkeit als Bedingung fürs Textverstehen. In: Alfa-Forum Vol. 76, S. 15–17.

Tarone, Elaine/Bigelow, Martha/Hansen, Kit (2009): Literacy and Second Language Oracy. Oxford, New York: Oxford University Press.

telc gmbH (Hg.) (2010): Handreichung. Prüferqualifizierung Deutsch A1-A2. Frankfurt a. M: telc gmbH.

telc/Goethe Institut (2017): Übungstest 1. Start Deutsch 1. Prüfungsvorbereitung. Online unter:
<https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/detail/telc-deutsch-a1.html#t=2> [27.01.2020].

The World Bank (2019): Literacy rate, adult total (% of people ages 15 and above) | Data. In: Data. Online unter:
<https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?contextual=default&end=2017&locations=ER-AF-SY-LK-TR-IQ-IR-SO-BA-VN&start=1975&view=chart> [27.01.2020].

Tracy-Ventura, Nicole et al. (2014): „Repeat as much as you can“. Elicited Imitation as a measure of oral proficiency in L2 French. In: Leclercq, Pascale/Hilton, Heather/Edmonds, Amanda (Hg.): Measuring L2 Proficiency: Perspectives from SLA. Bristol: Multilingual Matters. S. 143–166.

Tranza, Efisia/Sunderland, Helen (2009): Literature Review on Acquisition and Development of Literacy in a Second Language. London: London South Bank University. Prepared for National Adult Literacy Agency (NALA). Online unter:
https://www.nala.ie/sites/default/files/publications/Acquisition%20and%20Development%20of%20Literacy%20as%20level%20-%20literacy%20review_1.pdf [27.01.2020].

- Trim, J. L. M. et al. (Hg.) (2009): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin: Langenscheidt.
- Tröster, Monika/Schrader, Josef (2016): Alphabetisierung, Grundbildung, Literalität: Begriffe, Konzepte, Perspektiven. In: Löffler, Cordula/Korfkamp, Jens (Hg.): Handbuch zur Alphabetisierung und Grundbildung Erwachsener. Münster, New York: Waxmann. S. 42–58.
- Tschirner, Erwin (2016): Lextra Deutsch als Fremdsprache nach Themen: Grund- und Aufbauwortschatz. Berlin: Cornelsen.
- Ulrich, Tanja (2016): Sprachtherapeutische Diagnostik mit dem Mottier-Test. Viele Normierungen und viele Fragezeichen. In: Logopädie Vol. 30, S. 22–29.
- UNHCR (2016): Global Trends. Forced Displacement in 2015. Geneva: UNHCR. Online unter: <http://www.unhcr.org/576408cd7.pdf> [27.01.2020].
- UNICEF (2013a): Eritrea Statistics. In: UNICEF. Online unter: https://www.unicef.org/infobycountry/eritrea_statistics.html [27.01.2020].
- UNICEF (2013b): Statistics Arab Republic. In: UNICEF. Online unter: https://www.unicef.org/infobycountry/syria_statistics.html [27.01.2020].
- van Tubergen, Frank (2010): Determinants of Second Language Proficiency among Refugees in the Netherlands. In: Oxford University Press Vol. 89, S. 515–534.
- Vinther, Thora (2002): Elicited imitation: a brief overview. In: International Journal of Applied Linguistics Vol. 12, S. 54–73.
- Wagner, Elvis (2014): Assessing Listening. In: Kunnan, Antony John (Hg.): The companion to language assessment. Volume 1: Abilities, Contexts, and Learners. Malden, MA, Oxford: Wiley Blackwell. S. 47–63.
- Weins, Cornelia (2010): Uni- und bivariate deskriptive Statistik. In: Wolf, Christof/Best, Henning (Hg.): Handbuch der sozialwissenschaftlichen Datenanalyse. Wiesbaden: VS Verlag für Sozialwissenschaften. S. 65–89.
- Weiss, Christel (2018): Kausalität (in der Statistik). Online unter: <https://www.umm.uni-heidelberg.de/inst/biom/lexikon/data/k012.html> [27.01.2020].
- Werner, Sybille (2014): Wortschatz und Textproduktion: wie lassen sich Schreibkompetenzen von Hauptschülern fördern? Stuttgart: ibidem-Verlag.
- Wild, Nicole/Fleck, Christine (2013): Neunormierung des Mottier-Tests für 5- bis 17-jährige Kinder mit Deutsch als Erst- oder als Zweitsprache. In: Praxis Sprache Vol. 3, S. 152–157.
- Wimmer, Heinz/Mayringer, Heinz (2002): Dysfluent Reading in the Absence of Spelling Difficulties: A Specific Disability in Regular Orthographies. In: Journal of Educational Psychology Vol. 94, S. 272–277.

Winkler, Steffi (2011): Sprachstandsermittlung und Lernfortschrittstest mal anders. Die „Elicited Imitation Task“ als ein Fenster zur lernersprachlichen Grammatik. In: Info DaF Vol. 38, S. 525–550.

Wu, Shu-Ling/Ortega, Lourdes (2013): Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. In: Foreign Language Annals Vol. 46, S. 680–704.

Young-Scholten, Martha (2018): What is our body of research? Palermo.

Zuppinger, Brigit et al. (2013): Kantonales Integrationsprogramm 2014–2017. Online unter: https://www.gef.be.ch/gef/de/index/migration_integration/migration_integration.assetref/dam/documents/GEF/SOA/de/Soziales/Publikationen/Migration/Grundlagen/Kantonales_Integrationsprogramm_KIP_2014_2017_de.pdf [27.01.2020].

Anhang

1. Beispiel eines Silbenprotokolls

	Datum	Übungsname	Inhalt, Beschreibung	Sozialform	Material	Zeit
<i>Beispiel</i>	16.07.2018	<i>Silben lesen</i>	<i>Silben lesen und Wörter bilden (z. B. So-fa)</i>	<i>PA</i>	<i>zerschnittene Wörter (von LP erstellt)</i>	<i>08:30–08:45</i>
		<i>Silbenpuzzle</i>	<i>Wörter bilden (Puzzle).</i>	<i>PA</i>	<i>Silbenpuzzle (von LP erstellt)</i>	<i>09:00–09:15</i>
		<i>Silbenpuzzle</i>	<i>Wörter in zwei Farben abschreiben (z. B. Au-to).</i>	<i>EA</i>	<i>Übungsheft</i>	<i>09:30–09:50</i>
		<i>Silbendiktat /Lernerfolgskontrolle</i>	<i>Diktat (gelernte Silben werden 2x wiederholt)</i>	<i>GA</i>	<i>Wortliste</i>	<i>10:15–10:30</i>

2. Unterrichtsbeispiele Silbenmethode

Übung/Unterrichtsvorschlag	Vorgang	Ziel
Silben in Wörtern klatschen oder Silbenschreiten (1)	Deutsche Wörter vorsprechen und Silben klatschen oder klopfen. Anfangs zweisilbige Wörter (Auto, Taxi, Kino, Kilo, Foto, Sofa, Juli,...)	Allgemeine Heranführung an Silben als Element der Sprache.
Silbentreppen (1)	Lesen von Silben von links nach rechts und in Treppen.	Syntheseübung
Silbenpuzzle (1), (2), (3)	Wörter bilden und sie in zwei Farben schreiben (1 Farbe pro Silbe).	Syntheseübung; sicheres lautgetreues Schreiben.
Silbenmosaik (1)	Buchstaben/Laute werden nach Möglichkeit in Silben eingeführt. Aus den Silben lassen sich auch bereits einfache Wörter zusammenschliessen.	Erkennen und Lautieren des elementaren Silbenmusters.
Silben würfeln (1) (2), (3)	Silben würfeln und zu Konsonanten schreiben (z. B. 1=a, 2=o, ...). Anschliessend Silben vorlesen und in einem weiteren Schritt Silben zusammenfügen und vorlesen.	Synthese von zwei Lauten zu einer Silbe.
Silbengitter (1), (2), (3)	Lernende sollen eigenständig zwei Buchstaben zusammenfügen – Zusammenziehen von Konsonanten und Vokalen.	Syntheseübung; sicheres lautgetreues Schreiben.
Silben in Sätzen oder Texten klatschen oder sprechen (1), (2), (3)	Sätze und Texte klatschen oder sprechen, rhythmisches Sprechen.	Bewusstseinsstärkung für die lautliche Ebene der Sprache.
Silbenzahl bestimmen (1), (2), (3)	Sobald Silbenklatschen gelingt, Anzahl der Silben in Wörter eigenständig bestimmen oder Wörter vorsprechen und der Lernende muss entscheiden, wo eine bestimmte Silbe vorkommt.	Silben erkennen.
Wortmemory (1), (2), (3)	Memory: Bild mit dem korrekten Wort verbinden.	Spielerisches Wiedererkennen eines Wortbildes in Druck- und Schreibschrift.
Silbenmosaik (2)	Buchstaben/Laute werden in Silben eingeführt. Aus den Silben lassen sich auch Wörter zusammenschliessen.	Erkennen und Lautieren des elementaren Silbenmusters.
Silben in Wörtern klatschen oder Silbenschreiten (2)	Deutsche Wörter vorsprechen und Silben klatschen oder klopfen. Zweisilbige Wörter mit komplexerer Lautstruktur (Mantel, Datum, Salat, Paket, Beruf, Frage, Blume, ...)	Allgemeine Heranführung an Silben als Element der Sprache.
Silben in Wörtern klatschen oder Silbenschreiten (3)	Deutsche Wörter vorsprechen und Silben klatschen oder klopfen. Längere Wörter (Krawatte, Lederjacke, ...).	Allgemeine Heranführung an Silben als Element der Sprache.
Silbenmosaik (3)	Buchstaben/Laute werden in Silben eingeführt. Aus den Silben lassen sich auch Wörter zusammenschliessen.	Erkennen und Lautieren des elementaren Silbenmusters.

(vgl. Feick et al. 2013, S. 66–68, 99–102; vgl. Albert et al. 2015, S. 49–52)

3. Hörverstehen Einstufungstest

Vorname: _____

Familienname: _____

Wie ist die richtige Reihenfolge? Bringen Sie die Bilder in die richtige Reihenfolge (1-5).
Sie hören jeden Dialog zweimal.

Beispiel

Dialog

0



Dialog



Dialog



Dialog



Dialog



Dialog



Welcher Dialog passt zu welchem Foto? (1–5)
Sie hören jeden Dialog einmal.

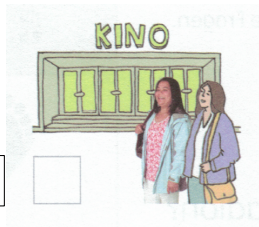
Beispiel

Dialog

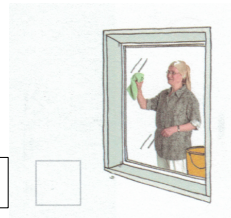
0



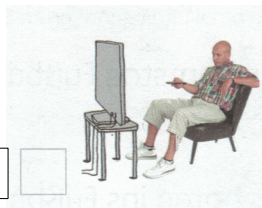
Dialog



Dialog



Dialog



Dialog



Dialog



Kreuzen Sie an: richtig oder falsch?
Sie hören den Text zweimal.

Anmeldung Einwohnerkontrolle Schaffhausen

Beispiel	<input checked="" type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>Scharapowa</u>
					Name
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>Tatjana</u>
					Vorname
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>Birkenstrasse 71</u>
					Strasse, Nr.
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>8200 Schaffhausen</u>
					PLZ, Ort
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>052 624 56 26</u>
					Telefon
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>Schweizerin</u>
					Nationalität
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>29 Jahre</u>
					Alter
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>verheiratet</u>
					Zivilstand
	<input type="checkbox"/>	richtig	<input type="checkbox"/>	falsch	<u>Informatikerin</u>
					Beruf
					<u>T. Scharapowa</u>
					Unterschrift

4. Hörverstehen Follow-up 1

Vorname: _____

Familienname: _____

Kreuzen Sie an: richtig oder falsch?
Sie hören den Dialog zweimal.



Herr Cimilodoro

Beispiel

Es ist halb acht.

richtig

☐

falsch

☒

1. Er ist fertig mit der Arbeit.

richtig

☐

falsch

☐

2. Er arbeitet schon 20 Jahre als Taxifahrer.

☐☐

3. Er spricht gern mit den Leuten.

☐☐

4. Er wartet nicht gern, das ist langweilig.

☐☐

5. Morgen arbeitet er von 14 Uhr bis 24 Uhr.

☐☐

6. Er fährt nicht gern in der Nacht.

☐☐

Welcher Dialog passt zu welchem Foto? (1–5)
 Sie hören jeden Dialog einmal.

Beispiel

Dialog



Dialog



Dialog



Dialog



Dialog



Dialog



Anmeldung Einwohnerkontrolle Schaffhausen
<u>Scharapowa</u>
Name
<u>Tatjana</u>
Vorname
<u>Birkenstrasse 71</u>
Strasse, Nr.
<u>8200 Schaffhausen</u>
PLZ, Ort
<u>052 624 56 26</u>
Telefon
<u>Schweizerin</u>
Nationalität
<u>29 Jahre</u>
Alter
<u>verheiratet</u>
Zivilstand
<u>Informatikerin</u>
Beruf
<u>T. Scharapowa</u>
Unterschrift

5. Hörverstehen Follow-up 2

Vorname: _____
Familienname: _____

Kreuzen Sie an: richtig oder falsch?
Sie hören den Dialog zweimal.

Beispiel	richtig	falsch
Trinken ist sehr wichtig für die Gesundheit.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	richtig	falsch
1. Der Körper braucht jeden Tag mindestens 1,5 Liter Flüssigkeit.	<input type="checkbox"/>	<input type="checkbox"/>
2. Früchte und Gemüse sind gesund. Sie haben viele Vitamine.	<input type="checkbox"/>	<input type="checkbox"/>
3. Reis, Mais, Teigwaren und Brot machen dick.	<input type="checkbox"/>	<input type="checkbox"/>
4. Auch Getreideprodukte haben Vitamine.	<input type="checkbox"/>	<input type="checkbox"/>
5. Der Mensch braucht jeden Tag eine Portion Fleisch oder Fisch.	<input type="checkbox"/>	<input type="checkbox"/>
6. Milch und Milchprodukte sind nur für Kinder wichtig.	<input type="checkbox"/>	<input type="checkbox"/>

Welcher Dialog passt zu welchem Foto? (1–5)
 Sie hören jeden Dialog einmal.


Beispiel

Dialog

0





Dialog


Dialog





Dialog



Dialog



Dialog



Kreuzen Sie an: richtig oder falsch?
Sie hören den Text zweimal.

		Anmeldung Einwohnerkontrolle Schaffhausen	
Beispiel	<input checked="" type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>Scharapowa</u>	Name
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>Tatjana</u>	Vorname
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>Birkenstrasse 71</u>	Strasse, Nr.
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>8200 Schaffhausen</u>	PLZ, Ort
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>052 624 56 26</u>	Telefon
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>Schweizerin</u>	Nationalität
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>29 Jahre</u>	Alter
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>verheiratet</u>	Zivilstand
	<input type="checkbox"/> richtig <input type="checkbox"/> falsch	<u>Informatikerin</u>	Beruf
		<u>T. Scharapowa</u>	Unterschrift

6. Diktatvorlage

Vorname: _____
Familienname: _____

Hören Sie und schreiben Sie die Wörter. Sie hören die Wörter viermal.

Beispiel:
1. <u>Ende</u>
2. _____

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

7. _____

8. _____

20. _____

21. _____

22. _____

Hören Sie und schreiben Sie die Sätze. Sie hören die Sätze fünfmal.

Beispiel:
1. <u>Das Internet ist schnell</u>

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

9. _____

10. _____

11. _____

12. _____

13. _____

14. _____

15. _____

16. _____

17. _____

18. _____

19. _____

7. _____

8. _____

7. Lehrerfragebogen

Fragebogen Lehrperson

Ich möchte Sie bitten, mir bei meinem Forschungsprojekt der Universität Fribourg zu helfen, indem Sie folgende Fragen ehrlich beantworten. Ziel dieser Studie ist es, die Gruppe der Zweitschriftlernenden zu untersuchen. Ich interessiere mich für Ihren Arbeits-, Weiterbildungs- und Erfahrungshintergrund. Die Studie dient ausschliesslich wissenschaftlichem Zwecken und die Einhaltung des Datenschutzes wird garantiert. Bitte beantworten Sie die Fragen korrekt. Ich danke Ihnen herzlichst für Ihre Hilfe.

1. Name:

2. Geschlecht:

<input type="checkbox"/>	<input type="checkbox"/>
Weiblich	Männlich

Alter:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unter 20 Jahren	20-29 Jahre	30-39 Jahre	40-49 Jahre	50-59 Jahre	60-65 Jahre	Über 65 Jahre

3. Nationalität:

4. Muttersprache:

<input type="checkbox"/>	<input type="checkbox"/>
Deutsch	Andere: <hr/>

5. Sprechen Sie Fremdsprache(n)? **Wenn ja**, welche und auf welchem Niveau:

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja
	Sprache: _____
	Niveau nach GER:
<input type="checkbox"/>	<input type="checkbox"/>
A1	A2
<input type="checkbox"/>	<input type="checkbox"/>
B1	B2
<input type="checkbox"/>	<input type="checkbox"/>
C1	C2

Sprache: _____

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A1	A2	B1	B2	C1	C2

Sprache: _____

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A1	A2	B1	B2	C1	C2

Sprache: _____

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A1	A2	B1	B2	C1	C2

6. Wann haben Sie Ihre letzte Fremdsprache gelernt:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Keine	Momentan	Vor 1-3 Jahren	Vor 4-6 Jahren	Vor 7-9 Jahren	Über 10 Jahre her

7. Sprechen Sie eine Fremdsprache mit einem anderen Schriftsystem? **Wenn ja**, welche, seit wann und auf welchem Niveau?

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja, seit: _____
	Auf welchem Niveau GER:
<input type="checkbox"/>	<input type="checkbox"/>
A1	A2
<input type="checkbox"/>	<input type="checkbox"/>
B1	B2
<input type="checkbox"/>	<input type="checkbox"/>
C1	C2

8. Seit wann arbeiten Sie bei [REDACTED]

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weniger als 1 Jahr	1 Jahr	2-3 Jahre	4-5 Jahre	5-10 Jahre	10-15 Jahre	Über 15 Jahre

9. Wie lange unterrichten Sie schon DaZ-Kurse:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weniger als 1 Jahr	1 Jahr	2-3 Jahre	4-5 Jahre	5-10 Jahre	10-15 Jahre	Über 15 Jahre

9.1. Wie viele Kursstunden sind das ungefähr:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
0 Stunden	1-100 Stunden	101-200 Stunden	201-300 Stunden	301-400 Stunden	401-500 Stunden	Über 500 Stunden

10. Seit wann unterrichten Sie in Alphabetisierungskursen:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weniger als 1 Jahr	1 Jahr	2-3 Jahre	4-5 Jahre	5-10 Jahre	10-15 Jahre	Über 15 Jahre

10.1. Wie viele Kursstunden sind das ungefähr:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
0 Stunden	1-100 Stunden	101-200 Stunden	201-300 Stunden	301-400 Stunden	401-500 Stunden	Über 500 Stunden

11. Seit wann unterrichten Sie MigrantInnen:

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weniger als 1 Jahr	1 Jahr	2-3 Jahre	4-5 Jahre	5-10 Jahre	10-15 Jahre	Über 15 Jahre

12. Haben Sie sonstige Sprachlehrerfahrungen? **Wenn ja**, was für welche:

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:

	Wie lange in Jahre:

13. Haben Sie ein SVEB-Zertifikat? **Wenn ja**, welches Modul (z.B. 1):

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:

14. Haben Sie ein Lehrerpapent? **Wenn ja**, geben Sie bitte an, was für eins:

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:

15. Haben Sie einen Abschluss auf Tertiärstufe (z.B. PH, Uni, ...)? **Wenn ja**, geben Sie bitte an, was für einen (Art und Fach):

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:

16. Haben Sie Weiterbildungen bzw. Zusatzqualifikationen im Bereich DaZ besucht? **Wenn ja**, geben Sie bitte an, was für welche:

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:

17. Haben Sie Weiterbildungen bzw. Zusatzqualifikationen im Bereich Alphabetisierung besucht? **Wenn ja**, geben Sie bitte genau an, was für welche:

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:
	<hr/>

18. Haben Sie eine telc oder Goethe-Prüferlizenz A1/A2?

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja:
	<hr/>

19. Haben Sie Unterrichtserfahrung auf Niveau A1 mit MigrantInnen?

<input type="checkbox"/>	<input type="checkbox"/>
Nein	Ja

20. Was ist Ihre Meinung zum deutschsprachigen Kurskonzept? Ist es Ihrer Ansicht nach möglich, innerhalb von 225 Kursstunden (à 60 Minuten) das Niveau A2 nach GER zu erreichen? (Begründen Sie bitte Ihre Antwort)

Die gesammelten Daten werden 100% anonym behandelt und dienen ausschliesslich wissenschaftlichen Zwecken für eine Doktorarbeit an der Universität Fribourg. Die Teilnahme an dieser Studie ist freiwillig.

Ich willige ein, dass meine Daten in anonymisierter Form benutzt werden können.

Bern, _____

Ort, Datum

Unterschrift

Sie sind nun am Ende des Fragebogens angekommen.

Ich danke Ihnen vielmals für Ihre Zeit.

8. Lernerfragebogen Deutsch

Vorname: _____

Familienname: _____

Fragebogen Deutsch

Ich möchte Sie bitten, diesen Fragebogen offen und ehrlich auszufüllen. Ich interessiere mich für Ihren Bildungshintergrund in Ihrem Heimatland. Das ist kein Test, das heisst es gibt keine richtigen oder falschen Antworten. Die Antworten werden 100% anonym behandelt und nicht an andere Personen weitergegeben. Vielen Dank.

1. Geschlecht:

<input type="checkbox"/>	<input type="checkbox"/>
weiblich	männlich

2. Geburtsjahr:

3. Heimatland:

3.1. Was ist Ihre Muttersprache?

4. Haben Sie Kinder in der Schweiz?

<input type="checkbox"/>	<input type="checkbox"/>
Ja	Nein

Wenn ja, wie alt sind sie?

5. Welche Sprache sprechen Sie zu Hause?

6. Wie ist Ihr allgemeiner Gesundheitszustand?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr gut	Ziemlich gut	Mittelmässig	Nicht gut	Gar nicht gut

7. Wie viele Schuljahre haben Sie in Ihrem Heimatland besucht?

8. Was ist Ihr höchster Schulabschluss? (z.B. Primarschule, Sekundarschule, Gymnasium, Universität)

9. Haben Sie eine Ausbildung in Ihrem Heimatland absolviert?

<input type="checkbox"/>	<input type="checkbox"/>
Ja	Nein

Wenn ja, was für eine?

10. Was war Ihre Arbeit in Ihrem Heimatland?

11. Sprechen Sie eine oder mehrere Fremdsprachen?

<input type="checkbox"/>	<input type="checkbox"/>
Ja	Nein

Wenn ja, was sprechen Sie?

12. Haben Sie schon einen Deutschkurs in der Schweiz besucht?

<input type="checkbox"/>	<input type="checkbox"/>
Ja	Nein

Wenn ja, Was für einen? (z.B. Alphabetisierung, A1-Kurs, A2-Kurs etc.)

13. Können Sie in Ihrer Muttersprache schreiben?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr gut	Ziemlich gut	Mittelmässig	Nicht gut	Gar nicht

14. Können Sie in Ihrer Muttersprache lesen?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr gut	Ziemlich gut	Mittelmässig	Nicht gut	Gar nicht

15. Können Sie in einer oder mehreren Fremdsprachen schreiben?

<input type="checkbox"/>	<input type="checkbox"/>
Ja	Nein

Wenn ja in welcher? _____

Und wie gut?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr gut	Ziemlich gut	Mittelmässig	Nicht gut	Gar nicht

Wie gut lesen Sie in dieser Fremdsprache?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr gut	Ziemlich gut	Mittelmässig	Nicht gut	Gar nicht

16. Wie oft brauchen Sie die deutsche Sprache in Ihrem Alltag?

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Täglich	4-5x/Woche	2-3x/Woche	1-2x/Woche	Gar nicht

17. Für was brauchen Sie die deutsche Sprache? (Mehrfachantworten möglich)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nachbarn und Freunde	Kinder	Arbeit	Arbeitssuche	Behörden

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Medien und Freizeit	Verkehr	Einkäufe	Post, Bank und Versicherungen	Gesundheit

<input type="checkbox"/>
Andere: _____

18. Bitte schreiben Sie eine kurze Geschichte über sich oder Ihre Familie (ca. 5 Sätze).

Die gesammelten Daten werden 100% anonym behandelt und dienen ausschliesslich wissenschaftlichen Zwecken für eine Doktorarbeit an der Universität Fribourg. Die Teilnahme an dieser Studie ist freiwillig.

Ich willige ein, dass alle meine geschriebenen und gesprochenen Daten in anonymisierter Form benutzt werden können.

Bern, _____
Ort, Datum

Unterschrift

Vielen Dank für Ihre Zeit.

Anhang, separat gebunden

Inhaltsverzeichnis

TEIL II: METHODOLOGIE	3
3. FORSCHUNGSGEGENSTAND	3
3.3.1 KURSEINTEILUNG	3
3.3.2 SAMPLE	5
4. TESTKONSTRUKTION	6
4.3.2 LERNERFRAGEBÖGEN	6
Tigrinya	6
Arabisch	11
Persisch	16
Tamil	22
Tibetisch	28
4.2.1 / 4.2.3 TESTERGEBNISSE ELICITED IMITATION / ORF	34
Einstufung	34
Follow-up 1 Institution P	61
Follow-up 1 Institution Q	74
Follow-up 2 Institution P	90
Follow-up 2 Institution Q	103
4.2.3.2 ORAL READING FLUENCY LESETEXT INKL. LESBARKEITSINDEX	119
5. DATENGRUNDLAGE	120
5.1 PILOTIERUNG/ERPROBUNG	120
Einstufung	120
Follow-up 1	120
Follow-up 2	121
5.2 PRÜFUNGSHANDREICHUNG	122
5.3 KODEBUCH	124
Berechnungen	127
Gestrichene Variablen	129
5.3.3 DUMMY-REFERENZ FÜR LINEARE REGRESSION	129
TEIL III: ERGEBNISSE	130
6. ANALYSE	130
6.1.2.1 INTERNE KONSISTENZ	130
Elicited Imitation	130
Einstufung	130
Follow-up 1	131
Follow-up 2	133
Kunstwörter	134
Einstufung	134
Follow-up 1	135
Follow-up 2	136
Hörverstehen	139
Einstufung	139
Follow-up 1	140
Follow-up 1	140
Diktat	142
Einstufung	142
Follow-up 1	145
Follow-up 1	148
6.1.2.2 RETEST-RELIABILITÄT	152
Elicited Imitation	152
Diktat	152
6.1.2.3 BEWERTER-RELIABILITÄT	154

Elicited Imitation	154
Rater 1	155
Rater 2	156
Kunstwörter	161
Rater 1	163
Rater 2	165
Diktat.....	167
Rater 1	170
Rater 2	170
6.2 ANALYSE DER TELC-PRÜFUNGSRESULTATE.....	178
6.2.1 <i>Deskriptive Analyse</i>	178
Institution	178
Kurslevel	179
Geschlecht	182
Alter	183
Erstsprache	185
Wortschatz.....	187
Formale Richtigkeit.....	188
Fremdsprache	190
Anzahl Schuljahre.....	192
Arbeit	193
Berufslehre	194
Gebrauch des Deutschen.....	195
Deutschkurs	197
Silbenmethode.....	198
6.2.2 <i>Die Treatments im Vergleich</i>	205
Hören	205
Lesen.....	206
Schreiben	208
Sprechen	210
6.3 ANALYSE DES GESAMTLERNERFOLGS BEI DEN FORSCHERGENERIERTEN TESTS.....	212
6.3.1 <i>Deskriptive Auswertung</i>	212
Institution	212
Kurslevel	213
Geschlecht	214
Alter	215
Erstsprache	216
Wortschatz.....	216
Formale Richtigkeit.....	218
Fremdsprache	219
Anzahl Schuljahre.....	220
Arbeit	221
Berufslehre	222
Gebrauch des Deutschen.....	223
Deutschkurs	224
Silbenmethode.....	224
6.3.2 <i>Die Treatments im Kursverlauf</i>	226
6.3.3 <i>Die Treatments bei den forschergenerierten Tests</i>	232
Elicited Imitation.....	232
Hörverstehen	234
ORF.....	236
Diktat	239
Gauss-Markov-Annahmen	242
6.4 ANALYSE DER KUNSTWÖRTER	257
TEIL IV: DISKUSSION	262
7. DISKUSSION DER EMPIRISCHEN ERGEBNISSE	262
7.3.3 <i>Regressionsmodelle mit Mottier-Test</i>	262
telc	262
Gesamtlernerfolg.....	264

Anhang, CD-ROM

Inhaltsverzeichnis

1. SPSS-Dateien

- a. SPSS-Datei: TN-Variablen
- b. SPSS-Datei: Rater

2. Audio-Dateien

- a. Einstufung: Elicited Imitation, Kunstwörter, Hörverstehen, Diktat
- b. Follow-up 1: Elicited Imitation, Kunstwörter, Hörverstehen, Diktat
- c. Follow-up 2: Elicited Imitation, Kunstwörter, Hörverstehen, Diktat

3. Testergebnisse Diktat