



# Evaluieren und Prüfen in DaF / DaZ

Christina Maria Ersch (Hg.)

Christina Maria Ersch (Hg.)  
Evaluieren und Prüfen in DaF / DaZ

Christina Maria Ersch / Marion Grein (Hg.)  
DaF / DaZ in Forschung und Lehre  
Band 2

Christina Maria Ersch (Hg.)

# Evaluieren und Prüfen in DaF / DaZ



CC-BY-NC-ND

ISBN 978-3-7329-9248-5

ISSN 2698-7740

DOI 10.26530/20.500.12657/43132

© Frank & Timme GmbH Verlag für wissenschaftliche Literatur  
Berlin 2021. Alle Rechte vorbehalten.

Das Werk einschließlich aller Teile ist urheberrechtlich geschützt.  
Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar.  
Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Herstellung durch Frank & Timme GmbH,  
Wittelsbacherstraße 27a, 10707 Berlin.

Printed in Germany.

Gedruckt auf säurefreiem, alterungsbeständigem Papier.

# Inhaltsverzeichnis

CHRISTINA MARIA ERSCH

Einführung ..... 7

## **EVALUIEREN**

CHRISTINA MARIA ERSCH

Motivation statt Isolation –

Evaluation der digitalen Lehrkräfteausbildung ..... 13

MARION GREIN

Evaluation der unterschiedlichen virtuellen Sprachkursformen ..... 49

ANN-KATRIN FIERUS

Learn to flirt by *Duolingo* „Du siehst aus wie meine nächste Freundin“ –

Die Evaluation von Sprachlern-Apps für Deutsch als Fremdsprache

in Bezug auf die Vermittlung interkultureller Kompetenz ..... 73

VIRGINIA SÁNCHEZ ANGUIX

Geeignet zum DaF/DaZ-Lehren?

Evaluation der Rolle sozialer und personaler Kompetenzen

in Eignungstests für DaF/DaZ-Studierende ..... 111

NINA JEHLE

Die medizinische Fachsprachenprüfung –

Eine Pilotstudie zur Evaluation durch die Teilnehmenden ..... 149

## **PRÜFEN**

NINA JEHLE

Fachsprachenprüfungen als Nachweis der Sprachkompetenzen?! –  
Ein Vergleich mit dem neu entwickelten  
*Sprachtest für ausländische Mediziner (SAM)* ..... 183

MARION GREIN

Prüferqualifizierung – Fokus auf Mündlichkeit ..... 229

JOSHUA ZIEGLER & MIRIAM RIEDINGER

Sprachliche Heterogenität und deren Einfluss auf das Hörverstehen  
in Sprachprüfungen – eine Vokalidentifikationsstudie mit  
Deutschlernenden und Deutschsprechenden ..... 273

Zu den Autorinnen und Autoren ..... 311

## Einführung

Mit dem vorliegenden Band erscheint der zweite der 2019 gestarteten Reihe *DaF/DaZ in Forschung und Lehre*, wieder getreu dem Motto: von Lehrenden für Lehrende. Unbedingt erwähnt werden muss, dass es sich auch bei diesem Band bei allen Autorinnen und Autoren wieder um Absolventinnen und Absolventen und/oder Lehrende der Johannes Gutenberg-Universität Mainz handelt. Dies verdeutlicht die Qualität, mit der Lehre und Forschung vor allem im Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache, einem vergleichsweise kleinen Studiengang seiner Art, betrieben wird. Die erste interdisziplinäre Zusammenarbeit mit dem Neurolinguistics Lab bestärkt die Vielfältigkeit, mit der hier geforscht wird. Vielfältig sind auch die Beiträge, die sich in zwei Themenschwerpunkte zusammenfassen lassen: Evaluieren und Prüfen.

Der erste Teil enthält Evaluationen von verschiedensten DaF/DaZ-Bereichen, wobei ein Fokus auf Formen und Möglichkeiten des digitalen Lernens liegt.

**Christina Maria Ersch** stellt eine erste Evaluation des digitalen Semesters im Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache vor und zeigt, wie die Problematik der sog. digitalen Distanz, der fehlende persönliche Kontakt bei virtueller Lehre, ausgeglichen werden kann und welchen Einfluss dies auf die Motivation zum Lernen hat. Sie verdeutlicht dies anhand des Seminars zur Interkulturellen Kommunikation, das im Sommersemester 2020 erstmalig mit dem *Inverted Classroom Modell* realisiert wurde.

**Marion Grein** greift dieses Modell auf, das eins der fünf aktuellen Typen des virtuellen Sprachunterrichts darstellt. Sie fokussiert dabei ebenfalls die bereits erwähnte digitale Distanz und überprüft mit einer kleinen Studie mit Lehrenden, inwieweit das Kursformat nach dem *Inverted Classroom Modell* durch zahlreiche soziale und kollaborative Tools tatsächlich erfolgreicher ist als andere Formen des virtuellen Sprachunterrichts.



**Ann-Katrin Fierus** vergleicht die drei größten Anbieter von Sprachlern-Apps für Deutsch als Fremdsprache *Babbel*, *Busuu* und *Duolingo*, die durch Ausgangsbeschränkungen und Lockdowns vieler Länder ab März 2020 einen deutlichen Zuwachs an Nutzer/innen verzeichnen konnten. Sie evaluiert, ob und wie interkulturelle Kompetenz mit digitalen Anwendungen aufgegriffen wird und inwieweit Lernende dadurch für kulturelle Unterschiede sensibilisiert werden können.

**Virginia Sánchez Anguix** stellt in ihrem Beitrag fest, dass die Relevanz der sozialen und personalen Kompetenzen in der Berufswelt allgemein anerkannt ist und für das Lehramt zumindest diskutiert wird, aber im DaF/DaZ-Bereich eine vergleichbare Eignungsdiskussion bisher nicht stattfindet. Sie evaluiert dabei die Rolle, die diese Kompetenzen bei Eignungstests für DaF/DaZ-Studierende an verschiedenen Universitäten spielt.

**Nina Jehle** präsentiert die erste Evaluation der fachärztekammerlichen Fachsprachenprüfung Medizin aus der Perspektive der Teilnehmenden. Die Ergebnisse ermöglichen einen Einblick auf die aktuelle Prüfungssituation sowie in die Erfahrungswelt der Prüfungsteilnehmenden.

Dies leitet über in den zweiten Teil des Bands, der sich dem Prüfen, genauer den Sprachprüfungen, widmet; ein Thema, das weiterhin ein desiderates Forschungsfeld darstellt. Dies liegt an der allgemeinen Problematik der Leistungsmessung im DaF/DaZ-Bereich bedingt durch die Faktorenkomplexion. Prüfungen und Tests müssen zuverlässige und valide Ergebnisse liefern. Zugleich sollen sie aber auch fair und vergleichbar sein (vgl. Rossa 2016, Schmelter 2014). Dies stellt TestentwicklerInnen immer vor neue Herausforderungen und wird gleichzeitig im Bereich der Sprachprüfungen viel zu selten kritisch untersucht. Studien wie bspw. von Wisniewski 2010, die sich mit der Bewertervariabilität und gleichzeitig mit der Validität der GER-Skalen beschäftigen, sind selten zu finden und zudem meist durch die geringe Stichprobe nicht generalisierbar. Die folgenden Beiträge stellen demnach einen weiteren Schritt zu Erforschung des Themas Prüfen dar, wobei besonders Validität und Fairness im Fokus stehen.

**Nina Jehle** vergleicht dabei die zwei gängigen Fachsprachenprüfungen im Bereich Medizin, die als Nachweis der von ausländischen Mediziner\*innen geforderten Fachsprachenkompetenzen gelten: die fachärztekammerliche Fach-

sprachenprüfung Medizin von Rheinland-Pfalz und die *telc Deutsch B2·C1 Medizin Fachsprachprüfung* und setzt sie dem neu entwickelten *Sprachtest für ausländische Mediziner* (kurz SAM) gegenüber, um diese hinsichtlich der Anforderung eines validen Nachweises zu evaluieren.

**Marion Grein** stellt in ihrem Beitrag die Frage, welche Voraussetzungen und Kompetenzen die Prüferinnen und Prüfer haben müssen, um mündliche Prüfungen abzunehmen, und wie überhaupt sichergestellt wird, dass die Prüfenden eine valide Einstufung vornehmen können. Sie vergleicht die Prüferqualifizierungen der drei Hochschulzugangssprachprüfungen *TestDaF*, *telc C1 Hochschule* und *DSH*.

**Joshua Ziegler und Miriam Riedinger** zeigen anhand einer Studie zur Identifikation von deutschen Langvokalen, dass mit der Heterogenität der Sprachhintergründe in Sprachkursen eine Chancenungleichheit bezüglich Hörverstehensaufgaben einhergeht. Sie plädieren für die Berücksichtigung der sprachlichen Hintergründe in der Prüfungsvorbereitung und -konzeption sowie die Festigung der phonologischen Grundkenntnisse auch auf hohem Sprachniveau.

An dieser Stelle möchte ich mich herzlich an die Mitarbeitenden des Verlags Astrid Matthes, Oliver Renner und nicht zuletzt Dr. Karin Timme wenden und ihnen für die hervorragende Unterstützung bei der Entwicklung dieses Bands bedanken. Ein noch größerer Dank gilt den Autorinnen und Autoren, die diesen Band erst möglich gemacht haben.

Die Reihe *DaF/Z in Forschung und Lehre* richtet sich gezielt nicht nur an Forschungsinteressierte in diesem Bereich, sondern an Lehrkräfte und Prüfende, denen die Beiträge in dieser Reihe konkrete Vorschläge für die Umsetzung in die Praxis geben sollen. Ergänzt werden muss dieses Mal die Gruppe der Testinstitutionen und damit der TestentwicklerInnen, denn gerade der letztgenannte Beitrag verdeutlicht die Evidenz, psycho- und neurolinguistische Befunde noch stärker in die Prüfungskonzeption mit einzubeziehen (ähnlich wie dies im Bereich Neurodidaktik bereits geschieht).

Christina Maria Ersch, Mainz im Oktober 2020

## Bibliografie

- ROSSA, H. (2016): „Prüfen und Testen“, in: BURWITZ-MELZER, E. et al. (Hrsg.): *Handbuch Fremdsprachenunterricht*, Tübingen: Francke: 399–402.
- SCHMELTER, L. (2014): „Gütekriterien“, in: DEMIRKAYA, S. et al. (Hrsg.): *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung*, Paderborn: Ferdinand Schöningh: 33–45.
- WISNIEWSKI, K. (2010): „Bewertervariabilität im Umgang mit GeR-Skalen. Ein- und Aussichten aus einem Sprachtestprojekt“, in: *Fremdsprache Deutsch* 3: 143–150.

EVALUIEREN



## Motivation statt Isolation – Evaluation der digitalen Lehrkräfteausbildung

### Einleitung

„Die soziale Eingebundenheit – durch die Lernumgebung – ist die zentrale Voraussetzung für Motivation und Freude am Lernen“ formuliert Grein (2020: o.S.) im Kontext des digitalen Sprachunterrichts. Doch wie kann motivierende Lehre funktionieren, wenn gemeinsame Präsenztreffen nicht stattfinden können? Welche Möglichkeiten gibt es, digital kollaboratives Arbeiten und damit den sozialen Austausch zu ermöglichen? Noch wichtiger wird dieser Aspekt bei den Lernenden bzw. Studierenden, die mit ihrem Studium erst begonnen haben bzw. beginnen werden und die dadurch wenig oder noch gar nicht sozial eingebunden sind. Und wie schafft man es, den individuellen Voraussetzungen der Studierenden, sei es technisch durch z.B. fehlende Hardware oder organisatorisch durch die Kita-Schließung während des Lockdowns, gerecht zu werden?

Eine Antwort könnte die Implementierung des *Inverted Classroom Modells* mit einem starken Fokus auf kollaborativem und sozialem Austausch sein, so wie es im Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache an der Johannes Gutenberg-Universität Mainz (JGU) im Sommersemester 2020 getestet wurde.

Das Ziel des vorliegenden Beitrags ist aufzuzeigen, mit welchen Mitteln bzw. auf welche Weise digitale Lehre zielführend eingesetzt werden kann, in welche Richtung sich Prioritäten verlagern müssen, welche Rolle Lehrende einnehmen müssen und welche Voraussetzungen bzw. welche Infrastruktur vorhanden sein muss, um Studierende zu motivieren und dadurch zu gewährleisten, dass sie trotz der ungewöhnlichen Situation dennoch die Chance haben, die Lernziele auch virtuell zu erreichen. Ferner wird erörtert, welche

Möglichkeiten es gibt, eine „Zoom-Fatigue“ zu vermeiden (vgl. dazu Heise 2020: o.S.), d.h. wie müssen synchrone Lehrangebote, die ohne Zweifel ein wichtiger Bestandteil der Lehre sind, aufgrund der kürzeren Aufmerksamkeitspanne in virtuellen Lernumgebungen (vgl. dazu die neurobiologische Studie von Chen et al. 2017 zu Aufmerksamkeitspanne beim E-Learning) gestaltet sein. Im Zentrum stehen dabei die Erfahrungen der Studierenden selbst, die durch die Corona-Pandemie im Hochschulbereich die größten Einschränkungen erfahren haben (vgl. Traus et al. 2020: 5f.).

Im ersten Kapitel erfolgt ein kurzer Überblick zur Ausgangssituation und zu den ersten Ergebnissen von Studierendenbefragungen. Im Anschluss erfolgt die Vorstellung zweier didaktischer Ansätze, die während des digitalen Semesters im Master DaF/DaZ eingesetzt wurden (Kapitel 1.2). Dabei wird vor allem auf die Seminare nach dem *Inverted Classroom Modell (ICM)* eingegangen und das Konzept und der Aufbau eines praxisintensiven Seminars zu Interkulturalität, das im Besonderen von Interaktion und Kommunikation lebt, skizziert (Kapitel 1.3). Daran anschließend werden im zweiten Kapitel die Ergebnisse einer qualitativen Vorstudie präsentiert, in der die Studierenden auf freiwilliger Basis allgemeine Aussagen zum digitalen Semester treffen sollten. Anhand dieser Aussagen wurde eine Befragung aller Studierenden generiert, um diese ersten Einschätzungen zu quantifizieren. Im Fokus der Online-Befragung stand die organisatorisch-technische Gestaltung der Seminare, die verschiedenen Aufgabenformen wie Forenbeiträge oder Erklärvideos und insbesondere die Möglichkeit zur persönlichen Kontaktaufnahme zu KommilitonInnen und die Betreuung bzw. Erreichbarkeit der Dozentinnen.

Im Fazit wird auf die Möglichkeiten eingegangen, welche die virtuelle Lehre ermöglicht. Klar ist, dass eine noch stärkere Sensibilisierung für die Bedürfnisse der Studierenden in Bezug auf ihre individuelle, in diesem Falle virtuelle Lernumgebung erfolgen und sich die Rolle der Lehrenden dabei langfristig ändern muss.

## 1 Die Ausgangslage und erste Ergebnisse aus Studien zum digitalen Sommersemester 2020

Welchen zahlreichen Herausforderungen sich die Bildungsinstitutionen in Deutschland aufgrund der Pandemie seit März 2020 stellen mussten, wurde in der Presse und sicherlich an jeder einzelnen Institution mehr als deutlich.

Die Folgen der Pandemie

bestätigte[n...] bereits lange bekannte strukturelle Problemlagen in Hochschulen oder manifestierte[n...] diese sogar. Insbesondere (materielle) Bildungsungleichheiten wurden durch die Voraussetzungen für das digitale Semester sichtbar und verstärkt. So setzt z.B. ein vollständig digitales Studium gewisse infrastrukturelle Ressourcen der Studierenden voraus, die nicht von allen gewährleistet werden können. (Traus et al. 2020: 6)

Fehlende Hardware bei Lehrkräften und Lernenden, unzureichende Software bzw. instabile Internetverbindungen und nicht zuletzt mangelnde digitale Kompetenzen auf beiden Seiten offenbarten, wie wenig Deutschland auf die digitale Lehre vorbereitet war. Doch die Forderung, digitale Inhalte in den Unterricht zu integrieren, ist nicht erst seit dem Ausbruch der Corona-Pandemie gegenwärtig (vgl. Grein 2020: o.S., vgl. dazu auch Grein & Strasser 2019). Dabei belegt das deutsche Bildungssystem in Bezug auf den Einsatz von digitalen Medien im Unterricht im internationalen Vergleich den hintersten Platz (vgl. Scheiter & Lachner 2020: 548), was durch den Aufbau einer bundesweiten digitalen Infrastruktur im Rahmen des Digitalpakts bis 2025 geändert werden soll (vgl. ebd., dazu auch DigitalPakt Schule 2019). Allein eine flächendeckende Umsetzung der Pläne war noch nicht erkennbar, ebenso im Hochschulbereich:

Es wurde viele Jahre lang über die digitale Lehre nur geredet, es wurden Workshops abgehalten, es wurden Strategiedebatten geführt – all das hat uns kaum geholfen, denn die Lehre an den deutschen Hochschulen befindet sich in großen Teilen immer noch im Tiefschlaf (Handke 2020: 34).



Wurde die Situation an den Hochschulen vergleichsweise wenig in der Presse thematisiert (anders als zur Situation an den Regelschulen), war diese dort nicht weniger prekär, zumal bis dato von einem eingeschränkten Regelbetrieb, wie es bspw. in Regelschulen oder Sprachschulen seit Ende der Sommerferien praktiziert wird, nicht gesprochen werden kann. Zwar gibt es Pläne zu hybriden Veranstaltungsformen, die Präsenzlehre in limitierter Form zulassen, aber der Großteil der universitären Lehre wird aufgrund der fehlenden festen Lerngruppen und nicht realisierbaren Kontaktlimitierung und -verfolgung bis ins Jahr 2021 noch (hauptsächlich) digital stattfinden müssen.

An vielen Universitäten wurde durch vereinzelte Erhebungen versucht, die Missstände, aber vor allem die Erfahrungen der MitarbeiterInnen und Studierenden aufzuzeigen, um ein möglichst genaues Bild bzw. einen Plan für das Wintersemester 2020/2021 bereitzustellen (siehe dazu die Studien bspw. der Universität Greifswald<sup>1</sup> oder an der TU Dresden<sup>2</sup>). Fokus war dabei häufig die organisatorische bzw. die technische Seite des digitalen Semesters. Eine deskriptive Untersuchung erfolgte an der Universität Hildesheim, die mit *Stu.diCo – Studieren digital in Zeiten von Corona* eine erste bundesweite Studie zu den Erfahrungen der Studierenden mit dem digitalen Semester durchgeführt hat (vgl. Traus et al. 2020).

Auch die JGU evaluierte das Sommersemester durch eine Befragung der Lehrenden und der Studierenden und schöpfte vor allem aus einzelnen Erfahrungsberichten (vgl. dazu CHANCEN JGU). Klar ist, dass die Studierenden das Erreichen der Studienziele etwas kritischer sahen, als die Lehrenden ihre Lehrziele (vgl. Videobotschaft *Wie ist das digitale Sommersemester gelaufen? Impulse der Vizepräsidenten für den Digitalen Tag der RMU 2020*<sup>3</sup>), ebenso unterscheiden sich die Präferenzen der Studierenden und Lehrenden in Bezug auf asynchrone bzw. synchrone Lehrangebote (vgl. ebd.), bei denen die

.....

- 1 Ergebnisse der qualitativen, unsystematischen Befragung unter Germanistik Studierenden, <https://t.co/qDKjPDzrCv?amp=1>, 24.09.2020.
- 2 Studierendenbefragung zur digitalen Lehre in Zeiten der Corona-Krise, [https://tu-dresden.de/gsw/phil/powi/polsys/ressourcen/dateien/forschung/umfrage-digitale-lehre/studierendenbefragung-digitale-lehre.pdf/at\\_download/file](https://tu-dresden.de/gsw/phil/powi/polsys/ressourcen/dateien/forschung/umfrage-digitale-lehre/studierendenbefragung-digitale-lehre.pdf/at_download/file), 24.09.2020.
- 3 Abrufbar unter: <https://www.rhein-main-universitaeten.de/news/wie-ist-das-digitale-sommersemester-gelaufen-impulse-der-vizepraesidenten-fuer-den-digitalen>, 23.11.2020.

Studierenden asynchrone Lehrangebote eher bevorzugen als die Lehrenden. Die Ausgangslage war für die verschiedenen Fachbereiche ebenso heterogen, was aber auch in „normalen“ Zeiten aufgrund der unterschiedlichen Studieninhalte und -strukturen der Fall ist. Daher ist auch verständlich, dass nicht ein „Masterplan“ für alle Fachbereiche erstellt werden kann, sondern durch den fächerübergreifenden Austausch situationsabhängige Handreichungen formuliert werden können.

Für den Masterstudiengang DaF/DaZ ergab sich eine vergleichsweise „luxuriöse“ Ausgangslage, die dazu führte, dass, neben technischen Neuerungen, der Blick auch auf den sozialen Aspekt des Lernens gelegt werden konnte. Dies war den Umständen geschuldet, dass der Studiengang schon lange digitale Komponenten in den Lehralltag integriert hat. Es sind für alle Seminare eigene Kurse auf der Lernplattform *Moodle* angelegt (seit dem Sommersemester werden die Kurse automatisch von *Jogustine*, dem Studieninformationsnetz der JGU, generiert, d.h. bei Anmeldung zu Seminaren werden Studierende automatisch in die entsprechenden *Moodle*-Kurse eingeschrieben (eine einmalige, separate Registrierung ist weiterhin notwendig).<sup>4</sup>

Die Lernplattform wurde und wird weiterhin für folgende Funktionen genutzt:

- Hochladen der Skripte, Handouts und Materialien für alle Seminare
- Inhalte des Grundlagenseminars in Form von digitalen Lernpaketen bestehend aus Skripten, Handouts, Erklärvideos
- Multimedia Modul mit dem Fokus auf digitalem Unterrichten
- Hochladen der Seminarprotokolle
- Kommunikationsmedium für Studierende (Gruppe *DaF Studis unter sich*), aber auch zur Weitergabe studienrelevanter Informationen z.B. aktuelle Anmeldefristen, geänderte Bibliothekszeiten etc. durch die Dozentinnen.

.....

4 Für den Herbst 2020 war der Start einer neuen *Moodle*-Version vorgesehen, was aber durch die Corona-Pandemie auf den Start des Sommersemesters vorgezogen wurde.

Alle genannten Funktionen dienten in der Vergangenheit zwar lediglich als Ergänzung zur Präsenzlehre, sorgten aber dafür, dass sowohl den Studierenden als auch den Lehrenden durch eine digitale Grundkompetenz der Umstieg auf die digitale Lehre zumindest technisch bzw. organisatorisch nicht allzu schwerfiel. Ebenso hilfreich war die Tatsache, dass es keine StudienanfängerInnen gab, da der Studienbeginn nur zum Wintersemester möglich ist. Dadurch konnte sich auf die didaktische Umsetzung einer 100% digitalen Lehre konzentriert werden, die aber mit den reinen *Moodle*-Kursen aufgrund der kompetenz- und kommunikationsorientierten Seminarformate nicht auskommen konnte.

Folgende Fragen standen daher bei den Überlegungen u.a. im Fokus:

- Wie kann digitale Lehre gestaltet werden, sodass die festgelegten Lernziele trotz Wegfall von Präsenzterminen erreicht werden?
- Wie muss die Lehre gestaltet werden, damit möglichst alle Studierenden gleichermaßen daran teilnehmen können?
- Wie kann ein Studiengang, der von Kommunikation und Austausch lebt, der handlungs- und produktionsorientiert ist und der kooperatives Arbeiten als zentrales Mittel für den Kompetenzerwerb ansieht, durch digitale Komponenten ergänzt werden?
- Wie kann die sog. digitale Distanz, also der fehlende persönliche Kontakt, ausgeglichen werden?

Folgt man der sozial-konstruktivistischen Lerntheorie, wird klar, dass gerade die beiden letzten Aspekte mit den ersteren zusammenhängen, da der Lernprozess als „eigenständige (Konstruktions-)Leistung des Lernenden im sozialen Kontext angesehen werden kann“ (Kergel & Heidkamp-Kergel 2020: 10). Noch deutlicher formuliert es Siemens (2004: 5, zit. in Kergel & Heidkamp-Kergel: 11) mit seinem konnektivistischen Ansatz, den er für das Lernen im digitalen Zeitalter als zentral erachtet: „Knowledge, that resides in a database, needs to be connected with the right people in the right context in order to be classified as learning“. Zwar ist in diesem Kontext Lernen selbstgesteuert, aber immer auch an eine Lerngemeinschaft gebunden (vgl. ebd.). Wie dies bei rein digitaler Lehre didaktisch umgesetzt werden kann, wird im Folgenden an zwei Modellen, die im Master DaF/DaZ implementiert wurden, exemplarisch vorgestellt.

## 1.1 Didaktische Umsetzung der digitalen Lehre

Bei der Umsetzung der im vorigen Kapitel kurz vorgestellten, konnektivistischen Lerntheorie gibt es vielfältige Möglichkeiten, die an die fachspezifischen Eigenheiten angepasst werden müssen. Die vorliegenden Modelle, die im Sommersemester im Master DaF/DaZ getestet wurden, lassen sich m.E. aber auch auf andere Fachbereiche übertragen, zumal, wie bereits erwähnt, zwei verschiedenen Ausrichtungen zur Anwendung kamen:

- a) Synchron: Virtuelle Seminare in Echtzeit
- b) Asynchron: *Inverted Classroom Modell*

Beide Modelle beinhalten die Nutzung der Lernplattform *Moodle* und virtuelle Präsenztreffen, die von Beginn an von zentraler Bedeutung waren. Zunächst wird die synchrone Variante skizziert, bevor etwas ausführlicher auf das *Inverted Classroom Modell* anhand eines konkreten Seminarbeispiels eingegangen wird.

a) Bei dieser vornehmlich synchronen Variante zu normalen Seminarzeiten wurden die Inhalte mit Hilfe des Videokonferenztools *Zoom* vermittelt. Wie auch in der klassischen Präsenzlehre vorgesehen, wurde der Input der Lehrenden durch Referate der Studierenden ergänzt; die Referierenden übernahmen dabei ebenfalls die Moderation der anschließenden Diskussionen. Die virtuellen Sitzungen wurden aufgezeichnet und auf der Lernplattform zum Download zur Verfügung gestellt. Dies war wesentlich, da die Anwesenheitspflicht ausgesetzt wurde, um die Studierenden, die bspw. Kinder oder Angehörige während des Lockdowns betreuen mussten, zeitlich zu entlasten. Partner- und Gruppenarbeiten wurden über die Breakout-Room-Funktion in *Zoom* realisiert. Die Lernplattform *Moodle* wurde ferner zum Hochladen von Handouts, Links und das Forum für allgemeine Ankündigungen durch die Dozentin genutzt. Ebenso der Link zu den *Zoom* Sitzungen wurde in den entsprechenden Kursen auf der Lernplattform veröffentlicht.

Der allgemeine Kontakt zu Lehrenden verlief über E-Mail, auch zeitlich festgelegte, virtuelle Sprechstunden wurden zunächst über *Zoom* organisiert,

später aber über Termine via *Microsoft Teams*. Ebenso das Masterkolloquium, in dem die Masterstudierenden über die Arbeit an ihren Abschlussarbeiten berichteten, lief über *Zoom*.

b) Beim zweiten Modell wurde das *Inverted Classroom Modell* (kurz *ICM*, auch bekannt als *Flipped Classroom*, vgl. dazu Bühner & Sommer 2020) angewandt bzw. auf die Bedarfe der Studierenden angepasst. Grundsätzlich erfolgt beim *ICM*

die Wissensaneignung durch multimediales Online-Material in individueller Arbeit (*Out-of-Class*), um die zeitlich nachgeschaltete Präsenzsitzung (*In-Class*) für die gemeinsame, vertiefende und interaktive Auseinandersetzung mit dem Gelernten nutzen zu können (RUBel: Was ist Inverted Classroom?).

Ziel dieses Modells ist die Möglichkeit zur Individualisierung von Lehr-Lernprozessen, indem Lernende durch digitale Elemente, wie z.B. Videos und Skripte, „zeitlich und örtlich flexibel die erforderlichen Grundlagen individuell [...] erarbeiten“ (RUBel: Zielsetzung und Zweck, vgl. dazu auch den Beitrag von Ortmann-Welp 2020) können. Dieses Modell, oft auch als Methode bezeichnet, wird an einigen Universitäten schon seit Jahren eingesetzt und ist demnach nicht neu. „Insgesamt fällt die Einschätzung des IC-Konzepts in der Literatur sehr durchwachsen aus – sowohl bei allgemeiner Betrachtung als auch bei der thematischen Eingrenzung auf Forschungsmethoden“ (Mertens et al. 2019: 18), doch für das rein digitale Semester bot es sich an, dieses Modell zu pilotieren.

Die Seminare konnten im Sommersemester aus bekannten Gründen nicht in Präsenz stattfinden, die *In-Class* fand demnach rein virtuell statt. Vorab bearbeiteten die Studierenden mit Hilfe angefertigter Skripte die Inhalte der Seminare. Sie bearbeiteten Aufgaben, meist schriftlich (also *Out-Of-Class*), und, statt *face-to-face*, fanden die Präsenzsitzungen virtuell über *MS Teams* statt. Die Skripte enthielten konkrete Arbeitsaufträge, die obligatorisch in vorgegebener Zeit bearbeitet werden mussten. Dabei handelte es sich meist nicht nur um „einfache“ Fragen, die z.B. in den Foren diskutiert werden sollten, sondern waren häufig mit kleineren Rechercheaufgaben kombiniert. Nicht

zuletzt sollten didaktische Überlegungen ergänzt werden. In den virtuellen Präsenzsitzungen über *MS Teams* wurden die Arbeitsaufträge bzw. „Produkte“ und ebenso offene Fragen diskutiert (*In-Class*).<sup>5</sup> Darüber hinaus wurde dieses Programm zu jeglicher Form von Kontaktaufnahme genutzt. Um die wöchentlichen Besprechungen zu den einzelnen Seminaren zu organisieren, wurden dort, parallel zu den *Moodle*-Kursen, eigene Teams erstellt, in denen die in die Seminare eingeschriebenen Studierenden Zutritt hatten.

Retrospektiv ist festzuhalten, dass damit ein virtueller Campus geschaffen wurde, da dort die Kommunikation in Echtzeit (im Gegensatz zu den Möglichkeiten auf *Moodle*) mit wenig Aufwand erfolgen konnte. Allen voran wurde durch Chatgruppen wie dem allgemeinen „Sprechstunden“ Chat eine Möglichkeit geboten, den Studierenden aus allen Semestern Informationen in Echtzeit zu vermitteln. Sie hatten hier die Möglichkeit, mit ihren KommilitonInnen und den Dozentinnen schneller Kontakt aufzunehmen, wie es über E-Mail der Fall ist. Es wurde außerdem ein Chat für die interne Kommunikation der Studierenden eingerichtet, zu dem die Dozentinnen keinen Zutritt hatten.

Des Weiteren wurde eine digitale Bibliothek eingerichtet, in der Links zu frei abrufbaren Büchern geteilt wurden oder wo Anfragen nach eingescannten Seiten aus vorherigen Semestern gestellt werden konnten, wenn diese nur in der Präsenzbibliothek verfügbar waren, aber durch die zeitweilige Schließung der Bibliotheken nicht genutzt werden konnten; ein Behelfsangebot, was, zusätzlich zur zeitweisen Öffnung der E-Librarys einiger großer Verlage, die problematische Literaturrecherche zumindest ansatzweise erleichtern konnte.

Wie eingangs erwähnt, wird exemplarisch am Seminar „Interkulturalität: psychologische und didaktische Aspekte“ gezeigt, wie das *ICM* konkret umgesetzt wurde.

## 1.2 *ICM* im Seminar „Interkulturalität: psychologische und didaktische Aspekte“

Bei diesem Seminar stehen folgende Lernziele im Fokus:

.....

- 5 Die Besprechungen wurden aus datenschutzrechtlichen Gründen nicht aufgezeichnet (Hauptgrund dafür war das direkte Hochladen der Aufzeichnung in *MS Stream*).

## Die Studierenden

- verfügen über Kompetenzen zur Reflexion der wissenschaftlichen Modelle der Kulturbeschreibung und zur interdisziplinären Analyse von interkulturellen Kontakten;
- haben ein methodisches Instrumentarium zum Kulturvergleich;
- leisten einen Praxis-Transfer durch Anwendung von Techniken des Interkulturellen Trainings (aus dem Modulhandbuch für den Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache, Version 3/2019).

Ganz im Sinne des *learning by doing* werden in diesem Seminar (normalerweise) Aktivitäten zu verschiedenen Bereichen der interkulturellen Kompetenzvermittlung (vgl. Schmidt-Borcherding et al. 2020: 245ff.) getestet. Dabei werden Themen, die jeweils in einem Grundlagenseminar zur Interkulturellen Kommunikation im Wintersemester angeboten werden, in die praktische Arbeit am Beispiel von Interkulturellen Trainings übertragen. Diese sind u.a. Culture Assimilator, Einsatz von Bildern und Filmen, interkulturelle Simulationen, Kulturrelativismus und Kulturfragebögen.

Die Inhalte und Aufgabenstellungen finden sich in dem bereits thematisierten Skript. Das Skript wurde im Verlauf des Semesters immer wieder aktualisiert, um die Erfahrungen mit bereits absolvierten Einheiten berücksichtigen zu können. Zur Vertiefung und Diskussionsanregung enthielt das Skript Arbeitsaufträge, die vornehmlich im Forum bearbeitet werden sollten.

Ein Beispiel zu einem Arbeitsauftrag aus dem Themenbereich Einsatz von Bildern in interkulturellen Trainings:



Lesen Sie noch einmal die formalen Merkmale von Bildern (S.39f.). Suchen Sie ein Bild, einen Bildausschnitt, Foto o.ä. zu einem auf Seite 40 genannten Thema und überlegen Sie ebenfalls, wie Sie das Bild in einem IKT einsetzen würden. Laden Sie Ihr Bild und Ihre Ideen ins „Forum Produktive Irritationen“ hoch, schauen Sie sich auch die Beiträge der Anderen an und kommentieren Sie!



Mit dieser Aufgabe sollten u.a. folgende Kompetenzen ausgebildet oder erweitert werden:

- Bildkompetenz (Formalia zu Auswahl und Rezeption von Bildern).
- Methodenkompetenz (Bilder mit dem Auftrag, zu „irritieren“, recherchieren, die außerdem frei abrufbar sind (z.B. auf *Pixabay* oder eigene Fotos) und zuvor präsentierten formalen Merkmalen entsprechen).
- Lehrkompetenzen: Didaktik/Methodik: Wie könnte man das Bild in einem interkulturellen Training in welchem Themenbereich einsetzen?
- Evaluation: Bewertung der Bilderauswahl und didaktischen Überlegungen der KommilitonInnen.

Im Folgenden ein Beispiel aus dem Forum mit einer Reaktion zu einem Bildvorschlag<sup>6</sup>:

*TN 1: Ich würde das folgende Bild als Einstieg in eine Stunde zum Thema „Feminismus weltweit“, bzw. „Rollenbilder“, nutzen.*



Quelle: <https://pixabay.com/de/photos/feministinnen-m%C3%A4rz-protest-frauen-4700823/>, Original in Farbe, 03.10.2020)

.....

6 Es handelt sich dabei um eine wörtliche, anonymisierte Übernahme aus dem Forum, die nicht korrigiert wurde.



Zunächst würde ich die TN das Foto objektiv beschreiben lassen (z.B. nur Frauen, alle mit ähnlicher Kleidung, mit erhobener Faust, rufend). Dann würde ich auf die Details eingehen lassen (z.B. alle ähnliche Kleidung). Falls den TN die Bedeutung noch nicht bekannt ist, würde ich Hypothesen aufstellen lassen, wofür beispielsweise das Stirnband und das Halstuch stehen könnte. Dann könnte man sich immer mehr vom Foto an sich entfernen und mehr auf damit zusammenhängende Aspekte kommen, – je nach Interessen und eigenen Erfahrungen der TN. So könnte man z.B. überleiten zu Machtverhältnissen zwischen Männern und Frauen in den unterschiedlichen Herkunftsländern, Ungleichheit und Gewalt gegenüber Frauen, aber auch (aktiver) Protest (es gibt ja viele Formen des Protests, also auch Alternativen zu „klassischen“ Demos).

**TN2: RE:** Danke [...] für dein Bild! Ich kann mir schon vorstellen, dass dein Bild im Unterricht eine spannende Diskussion aufwerfen würde. Da das Bild eine sehr starke Botschaft ausübt (mit schwarzen Kleidungen hätte ich beinahe geraten, es handelt sich um eine Demo von AfD in Leipzig, weil ich sie unglücklicherweise einmal zufällig miterlebte ...), finde ich dein sensibles Vorgehen zum Thema recht angemessen. Ich stimme dir zu, dass es ein emotionales Thema und manche Frauen einfach nicht auf einmal ihren Herz öffnen. Im Skript Seite 30 steht es zum Thema Heirat, dass es für muslimen Frauen sogar ein Tabu ist, über Ehemann mit Bekannten zu sprechen.

Da es sich um ein für die TN heikles und u.U. auch emotionales Thema handeln kann, sollte man dabei die Fragen zunächst möglichst allgemein stellen und insgesamt äußerst sensibel vorgehen und z.B. die Fragen an die jeweiligen Reaktionen der TN im Verlauf anpassen etc.

**TN3: RE:** Das Foto hat im wahrsten Sinne des Wortes Kraft! Vielen Dank! Es gibt einen tollen Bildband einer rumänischen Fotografin „The Atlas of Beauty“. Sie fotografiert Frauen auf der ganzen Welt und es geht weniger um tatsächliche Schönheit, sondern Frauen auf der ganzen Welt bekommen ein Gesicht. Hier die Homepage <https://theatlasofbeauty.com/> (Ich persönlich folge ihr schon sehr lange auf Facebook und Instagram).

Die Reaktionen zeigen, welche Position das Mitteilen von eigenen Erfahrungen einnimmt. Außerdem wird die Rezeption des Bildes in Ländern angesprochen, in denen feministische Proteste ein Tabuthema sein können, was eine direkte Bezugnahme auf den theoretischen Input des Skripts darstellt. Durch die Verknüpfung der Theorie mit den eigenen Erfahrungen werden Empathiefähigkeit und eine realistische Selbsteinschätzung gefördert, zwei Aspekte, die in der interkulturellen Bildung von zentraler Bedeutung sind (vgl. dazu das Fünf-Schubladen-Modell interkultureller Kompetenz von Leenen 2019: 143f.). Eher lange, reflektierte Beiträge wie diese waren symptomatisch für die Diskussionen in den Foren, kürzere Reaktionen waren eher selten.

Auf Inhalte und weitere Fragen wurde dann in den wöchentlichen *MS Teams*-Besprechungen eingegangen, außerdem wurden besonders gelungene Ergebnisse aus der Recherche oder spannende Diskussionen durch die Lehrenden aufgegriffen und direktes Feedback gegeben. Weiteres Feedback zu den Leistungen wurde schriftlich in den Foren direkt oder per E-Mail gegeben.

Ein weiterer Baustein im *ICM* waren Erklärvideos (vgl. Bühner & Sommer 2020: 130), die aus zeitlichen Gründen erst im Laufe des Semesters ergänzt wurden. Eingeleitet wurden die Seminare durch Einführungsvideos, in denen Lernziele und Aufbau der Seminare erklärt wurden. Weitere Videos behandelten weniger konkreten Input, sondern fokussierten den Aufbau der Studienleistungen (im obigen Falle die Vorgaben zur Konzeption eines eigenen interkulturellen Trainings). Diese wurden zum einen durch das Screencast Programm *Panopto*, im weiteren Verlauf aber mit Hilfe des webbasierten Video Tool *Powtoon* erstellt, das eine große Bandbreite an Gestaltungsmöglichkeiten z.B. Animationen und Musik bietet. Weitere Erklärvideos wurden durch zusätzliche Hilfskräfte erstellt; die Themen waren u.a. eine Reihe zum wissenschaftlichen Schreiben (Recherche – Strukturieren und Roter Faden – Themenfindung – Richtiges Zitieren).<sup>7</sup>

Die obigen Ausführungen geben einen Eindruck über den Aufbau lediglich eines Seminars in der *ICM*. Es steht aber exemplarisch für den Aufbau aller

.....

7 Für das Wintersemester 2020/2021 werden ebenfalls Erklärvideos als Unterstützung zur Betreuung der StudienanfängerInnen und in der Vermittlung des eigentlichen Seminarinhaltes integriert.

Seminare der Kategorie b. Im Hinblick auf das ebenfalls digital stattfindende Wintersemester 2020/2021 und zur Qualitätssicherung der konzipierten virtuellen Lernszenarien wurde eine Evaluation der Umsetzung notwendig, die sich nicht auf einzelne Seminare, sondern auf das komplette Studienangebot des Masters DaF/DaZ bezog. Der Aufbau und relevante Auszüge aus den Ergebnissen werden im nachfolgenden Kapitel vorgestellt.

## 2 Evaluation der digitalen Lehre im Sommersemester 2020

Die Studie zur Evaluation des digitalen Semesters im Masterstudiengang DaF/DaZ setzte sich aus einem qualitativen Erfahrungsaustausch und einer quantitativen Online-Umfrage zusammen. Die Zielgruppe waren die Studierenden aller Fachsemesters des Masterstudiengangs DaF/DaZ. In der Vorstudie wurden allgemeine Erfahrungen zu den beiden in Kapitel 1 beschriebenen Formen der virtuellen Lehre bzw. zu den verschiedenen didaktischen Ansätzen (a und b) erfragt, die als freie Kommentare, anonym und online über das digitale Pinnwand Tool *Padlet* verfasst werden konnten. Als Beispiele seien hier nur ein paar Aussagen aufgeführt, die zeigen, wie heterogen das Semester empfunden wurde, und die eine Quantifizierung durch eine nachfolgende Online-Umfrage notwendig machten:

1. *Ich finde das Onlinesemester nicht schlecht. Ich habe das Gefühl in DaF genau so viel zu lernen wie in Präsenzsitzungen. In anderen Studiengängen ist das leider nicht der Fall. Es wurde sich wirklich die Mühe gemacht, alles anzubieten, was möglich war, und das funktioniert auch wirklich gut. Manchmal treten zwar technische Probleme auf, aber genau so kann auch mal die Straßenbahn liegen bleiben. Ich beteilige mich genau so wie in Präsenz.*
2. *Mir persönlich fällt es schwer zuhause in den Arbeitsmodus zu kommen. Nichtsdestotrotz finde ich die Alternative mit dem Forum in Moodle gut, da hier jeder zu Wort kommen kann und niemand aufgrund evtl. Schüchternheit stumm bleibt. Die Diskussionen in dem*

*Forum befürworte ich somit. Allerdings bin ich kein Fan von Aufgaben, wo wir nur etwas zusammenfassen sollen. Hierbei ist bei mir der Lerneffekt eher gering, bzw. ist es eher eine zeitliche Beschäftigung. Ich kann verstehen, dass es Aufgaben geben muss, vor allem wenn Präsenz nicht verlangt ist, aber hierbei vielleicht mehr Forumsarbeit und weniger Zusammenfassen etc.*

3. *Wenn unsere Dozentinnen wie in diesem Semester immer online erreichbar und ansprechbar sind, kann ich mir das digitale Wintersemester vorstellen. Es fehlt mir nur der live Input [...].*
4. *Wie ich schon erwähnt habe, benehme ich mich ziemlich zurückhaltend und schüchtern im Unterricht, weil ich wahrscheinlich etwas unsicher mit der Sprache bin und habe Angst, Fehler zu machen. Bei der digitalen Form mache ich auch bestimmt noch viele schriftliche Fehler. Trotzdem bei Forumbeiträge habe ich die Möglichkeit, mehr Zeit nachzudenken und habe den Eindruck, dass ich meine Gedanken besser formulieren kann.*

Überraschend war Aussage 1, die keinen Unterschied in dem Erreichen der Lernziele zwischen Präsenz- und Online-Lehre feststellen konnte, was auch in der dritten Aussage anklingt, allerdings mit dem Hinweis auf die Erreichbarkeit der Lehrenden und dem Fehlen des Live-Inputs. Aussage 2 erwähnt gleich mehrere Aspekte, wie die fehlende Selbstlernkompetenz, die positiven Erfahrungen mit den Forendiskussionen, die dazu noch für introvertierte Lernende neue Chancen der Partizipation bietet, was sich mit dem Inhalt der vierten Aussage deckt. Kritik wird an Aufgaben zur Zusammenfassung einschlägiger Forschungsbeiträge geäußert.<sup>8</sup>

Diese Aussagen waren u.a. die Basis für eine quantitative Online-Umfrage, die nach Ende der Vorlesungszeit (14.07.–24.07.2020) durchgeführt wurde. Es nahmen insgesamt 46 Studierende aus allen Fachsemestern teil; es konnten alle ausgewertet werden, wobei nicht alle Studierenden noch aktiv an Seminaren teilnahmen, weshalb der Gesamterhebungsumfang je nach Frage schwankt. Sowohl die Erarbeitung als auch die Erfassung und Auswertung der Daten

.....

8 Diese stellen neben der Recherche und Didaktisierung eine weitere Aufgabenform dar.

erfolgte über die Umfrage-Software *Evasys*<sup>9</sup>, die besonders für Evaluation von Lehre und Seminaren geeignet ist und u.a. von vielen Universitäten genutzt wird (u.a. Johannes Gutenberg-Universität Mainz, Universität Bamberg, Universität Flensburg). Die Umfrage erfasste keine Angaben zur Person, um die Anonymität in einem überschaubaren Studiengang (ca. 80 Studierende) so weit wie möglich zu gewährleisten. Lediglich die Unterscheidung zwischen den Doppelstudierenden (DS), also den Studierenden, die neben dem Master DaF/DaZ z.B. noch einen weiteren Master studieren, und den reinen DaF/DaZ-Studierenden (im Folgenden als Einfachstudierende (ES) bezeichnet) wurde erfragt. Dadurch sollte eine Differenzierung zu Erfahrungen in anderen Studiengängen und vor allem die Angaben zum Workload der einzelnen Seminare gewährleistet werden.

Die Studierenden wurden zu den o.a. Aspekten befragt, wobei die Stichprobe aus aktiv Studierenden aus dem 2. und 4. Semester als auch aus Studierenden bestand, die zwar nicht mehr an allen Seminaren teilgenommen haben, aber durch die Arbeit an der Masterarbeit oder an anderen Studienleistungen ebenso von der veränderten Studiensituation betroffen waren.

Die Umfrage enthielt positive, aber auch teilweise negative Aussagen, die mit Hilfe von Likert Skalen bewertet wurden. Es wurde bei der Erstellung des Fragebogens vornehmlich eine gerade Anzahl von Antwortmöglichkeiten gewählt (sechs Stufen), bei Aussagen, die den Vergleich betrafen, wurden 5-stufige Antwortmöglichkeiten gewählt, damit eine Stagnation ebenfalls abgebildet werden konnte. Die Arbeit mit Likert Skalen ermöglichte eine schnelle Beantwortungszeit von 10–12 Minuten. Ebenso wurden zu verschiedenen Aspekten freie Kommentare ermöglicht, die sich im Nachhinein als sehr aufschlussreich herausstellten und in diesem Beitrag ebenso auszugsweise in der Analyse der Ergebnisse inkludiert werden.

Das Ziel der Umfrage war, ein Meinungsbild zur Umsetzung der digitalen Lehre mit Fokus auf folgenden Aspekten einzuholen:

1. Allgemeine Zufriedenheit mit der digitalen Lehre
2. Zuwachs allgemeiner Kompetenzen inkl. digitaler Kompetenzen

.....

<sup>9</sup> Informationen zur Software unter: <https://www.evasys.de/evasys.html>, 28.07.2020.

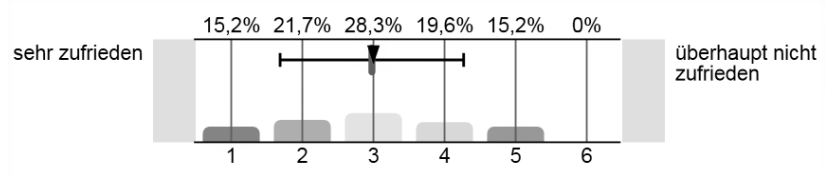
3. Kompetenzerwerb durch Referate
4. Rezeption der neu gestalteten Erklärvideos
5. Austausch und Diskussionen in den Foren
6. Bewertung der technischen Gestaltung der Lehre (beide Konzepte)
7. Bewertung der Seminare der Dozentinnen
8. Persönlicher Kontakt zu KommilitonInnen und Dozentinnen
9. Zufriedenheit mit der Erreichbarkeit bzw. der Betreuungssituation durch die Dozentinnen.

Im Folgenden wird auf die wichtigsten Ergebnisse dieser Umfrage eingegangen; Aspekte 3 und 7 werden nicht weiter erläutert, da sie sehr ortsbezogen bzw. studienfachbezogen sind und sich daher nicht generalisieren lassen.

## 2.1 Allgemeine Zufriedenheit mit der digitalen Lehre

Zu Beginn wurde die allgemeine Zufriedenheit mit der digitalen Lehre in Abgrenzung zur Zufriedenheit der digitalen Lehre im Masterstudiengang DaF/DaZ bewertet:

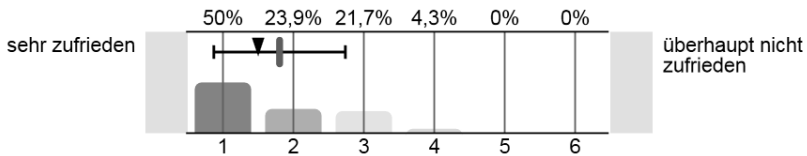
Wie zufrieden sind Sie allgemein mit der digitalen Lehre im Sommersemester 2020? (n= 46)



Wie zufrieden sind Sie allgemein mit der digitalen Lehre im Masterstudiengang DaF/DaZ?<sup>10</sup> (n=46)

.....

10 Aus ökonomischen Gründen wurde die Bezeichnung des Studiengangs abgekürzt.



Festzuhalten ist hier, dass die allgemeine Zufriedenheit mit der digitalen Lehre geringer ist (bei einem Mittelwert von 3) als bei der digitalen Lehre im Masterstudiengang (Mittelwert 1,8). Allerdings ist die Streuung breit. Der Unterschied bei den DS, denen man durch ihre Erfahrungen mit einem zweiten Fach einen Vergleich zu diesem eher nachsagen kann, fällt mit den Mittelwerten 3,2 zu 1,6 noch deutlicher aus. In einem freien Kommentar wird das verdeutlicht:

*Im Vergleich zu den M.[...] Seminaren empfand ich das digitale DaF-Studium als gut strukturiert und sehr lehrreich. [...] Insgesamt habe ich mich in diesem Studiengang weniger alleingelassen gefühlt wie in anderen M.[...]. Veranstaltungen.*

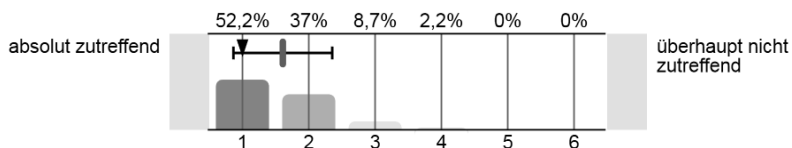
Dies bestätigt die positiven Erfahrungen, die in der Vorstudie schon deutlich wurden; das Thema Isolation bzw. im Umkehrschluss die vermehrte Erreichbarkeit wurde von den Studierenden unabhängig voneinander mehrfach erwähnt, ist demnach also etwas, was zur positiven Evaluation bzw. einer positiven digitalen Lernumgebung beitragen kann.

## 2.2 Zuwachs allgemeiner Kompetenzen

Auf Basis der Aussagen in der Vorstudie wurde nach der persönlichen Einschätzung über den Zuwachs einzelner Kompetenzen gefragt: Recherchekompetenz, Schreibkompetenz (bzw. Mediation: schriftliches Zusammenfassen) und digitale Kompetenz.

Wie zu erwarten, wurde der Zuwachs an digitaler Kompetenz am stärksten eingeschätzt:

### Ich habe meine digitale Kompetenz erweitert. (n=46)



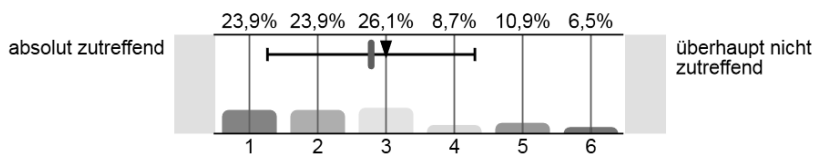
Mit einem Mittelwert von 1,6 ist das Gefühl, mit Digitalem besser umgehen zu können, deutlich erkennbar. Hierbei gibt es keine nennenswerten Unterschiede zwischen beiden Gruppen. Als Ergänzung wird hier eine Einschätzung einer Studentin mit offensichtlicher Unterrichtserfahrung ergänzt, die sich sicherlich sowohl auf Lernende, als auch auf viele Lehrende übertragen lässt:

*Vor Corona hatte ich mich ein wenig gegen digitales Unterrichten gewehrt. Jetzt musste ich mich damit auseinandersetzen und kann dem Ganzen doch etwas Positives abgewinnen.*

Dies ist m.E. einer der positivsten Effekte des Lockdowns, der hoffentlich einen Wandel hin zu einer schnelleren und flächendeckenden Digitalisierung führt. Dies bleibt jedoch abzuwarten und wird an dieser Stelle nicht weiter ausgeführt.

Bei der Erweiterung der Recherchekompetenzen und der Schreibkompetenz in Bezug auf schriftliche Zusammenfassungen zeigt sich ein eher uneinheitliches Bild:

### Ich habe meine Recherchekompetenz (d.h. Internetrecherche, Literaturrecherche) verbessert. (n=46)



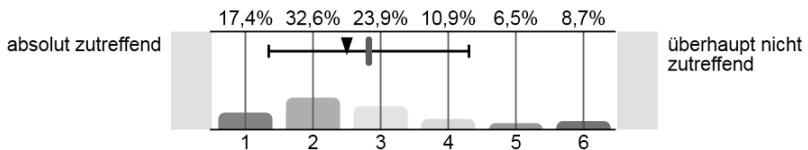


Doppelstudierende bewerten den Zuwachs dieser Kompetenz mit einem Mittelwert von 3,1 (und einer Standardabweichung von 1,5) etwas geringer als die reinen DaF-Studierenden (Mittelwert 2,5, Standardabweichung ebenfalls 1,5).

Eine Evaluation dieser Kompetenz wurde gezielt gefragt, da die Recherchekompetenz und im speziellen die Internetrecherche durch die zeitweilige Schließung der Präsenzbibliotheken zwangsweise ausgeweitet werden musste. Im *ICM* gab es darüber hinaus einige Aufgaben, in denen gezielt z.B. Bilder (s. Kapitel 1.3) oder weiterführende Literatur gesucht werden sollte, die es in der „herkömmlichen“ Variante in dieser Form nicht gab.

Ebenso wurde nach der schriftlichen Zusammenfassung gefragt, da dies auch eine Aufgabenform darstellte, um die Inhalte des Skripts zu vertiefen bzw. um Abstracts zu verfassen. Ferner wurde die Zusammenfassung mehrfach in der Vorstudie erwähnt, weshalb auch dies quantifiziert werden musste. Die durch die Vorstudie eher kritischen Aussagen bezüglich dieser Aufgaben konnte die Online-Umfrage nicht zu 100% verifizieren:

### Ich habe die Kompetenz des schriftlichen Zusammenfassens in diesem Semester verbessert. (n=46)



In den Kommentaren wurden Zusammenfassungen dreimal als positiver und gewinnbringender Aufgabentyp erwähnt.

Daraus resultiert zum einen, dass das Lernziel dieser Aufgabe noch deutlicher gemacht werden muss, damit klar wird, dass diese Aufgabe nicht nur dazu dient, den Workload zu füllen, sondern dass dies eine Kompetenz ist, die für das wissenschaftliche Arbeiten nötig ist und zu den Kompetenzen gehört, die Studierende im Laufe ihres Studiums erwerben müssen. Bei der Konzeption von Aufgaben für die *ICM* sollte aber zum anderen darauf geachtet werden, wie zielführend die Zusammenfassungsaufgaben für bestimmte Inhalte sind

und sie möglichst mit anderen Aufgabenformen abzuwechseln, damit diese nicht demotivierend wirken.

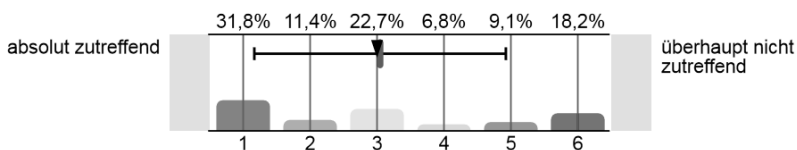
Neben den ausgewählten Kompetenzen wurde an dieser Stelle auch die Einschätzung über das Verhältnis zum allgemeinen Lerneffekt im Vergleich zum Präsenzseminar thematisiert. Ausschlaggebend waren folgende, eher überraschende Aussagen aus der Vorstudie:

- *Soweit ich hier und von anderen Kommilitonen erfahren habe, konnten sie miteinander viel mehr kommunizieren und Meinungen austauschen und dadurch noch effektiver lernen als in Präsenz-Veranstaltungen.*
- *Für das erste rein digitale Semester habe ich relativ gute Lernerfahrungen machen können. Ich habe mich viel mehr aktiv mit den Inhalten auseinandergesetzt. [...] Vor allem die (persönlichen) Beiträge in den Diskussionen haben viel bei mir angeregt und einen zusätzlichen Lerneffekt gehabt, da ich selbst auch erst einmal Zeit brauche, um meine Meinung mitzuteilen und im Seminar deshalb oft ruhiger bin.*

Hieraus ergab sich die Hypothese, dass das digitale Semester dazu führt, sich intensiver mit den Inhalten auseinanderzusetzen und dadurch einen stärkeren Lerneffekt zu erzeugen. Diese würde allen Annahmen und Bedenken, die gegenüber dem digitalen Lernen angebracht werden (vgl. dazu das Vorwort zur 1. Auflage und die Hinweise zu den zwei weiteren Auflagen von Handke 2020: 3ff.), endgültig widersprechen.

Dieser Effekt konnte ebenfalls nicht eindeutig verifiziert werden:

**Ich habe das Gefühl, durch die virtuelle Lehre im Sommersemester 2020 mehr gelernt zu haben, als ich es in regulären Präsenzseminaren getan hätte. (n=44)**



Es zeigt eine starke Abweichung (von 1,9) und dadurch ein heterogenes Bild. Doppelstudierende können dieses Gefühl noch weniger bestätigen als die reinen DaF-Studierenden.

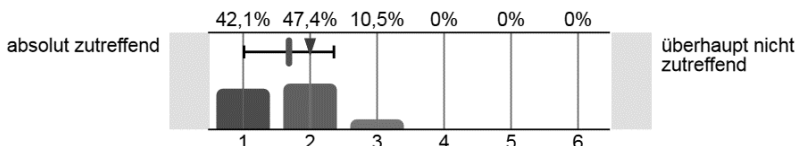
### 2.3 Evaluation der Erklärvideos

Bei der Auswertung der Ergebnisse in Bezug auf die Integration von Erklärvideos kann zwischen der reinen Rezeption und der eigenen Produktion unterschieden werden. Hierbei gab es keine signifikanten Unterschiede der beiden Gruppen, weshalb diese nicht separat beschrieben werden.

Erstere zeigte, dass die Rezeption der Videos von 67,7% (n= 31, mw 1,4, Std.Abw. 0,7) der Studierenden als absolut hilfreich und von 50 % (n=32, mw 1,7, Std.Abw. 0.8) als absolut motivierend bewertet wurden. Ebenso stark ist der Wunsch nach mehr Erklärvideos zu allgemeinen Themen wie Studienorganisation (wie sie vornehmlich im Sommersemester erstellt wurden), aber auch zur inhaltlichen Unterstützung innerhalb eines Seminars.

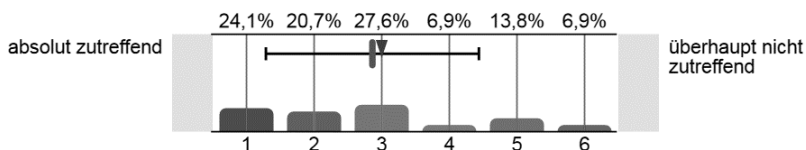
Zugleich wurde die Produktion von eigenen Videos positiv evaluiert. Die Studierenden konnten in einigen Seminaren wählen, ob sie als Studienleistung ein Referat halten oder ein Erklärvideo erstellen. Bemerkenswert waren die Einschätzungen, das eigene Fachwissen, die Präsentationskompetenz und vor allem die digitale Kompetenz durch die Produktion eigener Videos erweitern zu können. Interessant ist dabei die Korrelation der Studierenden, die sowohl ein Referat gehalten als auch ein Video produziert haben.

#### Ich habe mein Fachwissen durch die Produktion von Erklärvideos erweitern können (n=19):



Im Vergleich dazu lässt sich dieser Trend bei den Referaten nicht erkennen:

### Ich habe mein Fachwissen durch mein Referat/meine Referate erweitern können (n=29)



Allerdings haben nur insgesamt vier Studierende beide Studienleistungen erbracht, weshalb in diese Richtung noch weitere Befragungen notwendig sind. Doch auch diese vier geben an, deutlich mehr durch die Erklärvideos als durch Referate gelernt zu haben.

Sicher fehlt die Moderationskompetenz, die eher durch Präsentationen erworben werden kann, aber für die Verinnerlichung und die Fähigkeit, Inhalte ansprechend und anschaulich präsentieren zu können, ist die Produktion dieser Videoform auch außerhalb von Corona sinnvoll (vgl. dazu die Ausführungen von Maxin 2018: 619 zu Videos als Unterrichtsmethode). Dies betrifft auch die noch eher weniger häufig untersuchte Erklärkompetenz als Teil der professionellen Handlungskompetenz von Lehrenden (vgl. Ernst 2017: 158), die definiert werden kann „als die Fähigkeit, die spezifischen Anforderungen, kommunikative Aufgaben und Probleme, die sich in der Erklärinteraktion stellen, zu bewältigen“ (ebd.: 161, vgl. dazu auch die Definition von „Erklären“ im schulischen Kontext von Maxin 2018: 613f.). Ebenso der Zuwachs dieser wurde durch die Videografie als gegeben bewertet.

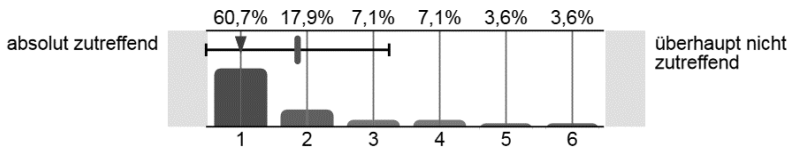
Sind die Ergebnisse der Umfrage ansonsten eher heterogen, lässt sich zusammenfassend sagen, dass die Evaluation der Erklärvideos auf allen Ebenen sehr positiv ausfällt und die Integration von Videos auf alle Fälle weiter vertieft werden sollte. Kritisiert wurde lediglich der erschwerte Zugang zu Beginn des Semesters (aus Kapazitätsgründen können Videos nicht auf den Servern der Lernplattform hochgeladen werden), der aber durch einen zentralen Ordner auf der Sync-and-Share-Plattform *Seafile*, zu dem alle Studierenden Zugang haben, verbessert wurde. Dies bedeutete allerdings ein weiteres Programm; wünschenswerter wäre eine Integration der Videos in bereits bekannte Tools (evtl. auch über *Youtube*).

Ferner spielten die Diskussionen in den Foren auf der Lernplattform eine wichtige Rolle bei der digitalen Lehre, weshalb im folgenden Kapitel näher darauf eingegangen wird.

## 2.4 Austausch und Diskussionen in den Foren

Gerade die Forenbeiträge wurden in der Vorstudie sehr häufig bewertet, der Tenor dabei war eher positiv, allerdings wurden die Länge und die Masse der zu lesenden und kommentierenden Beiträge kritischer gesehen. Die Ergebnisse der Umfrage zeigen allerdings, dass das Lesen der Forenbeiträge als hilfreich erachtet wurde (beide Gruppen). Erstaunlich sind die Ergebnisse, die die Tatsache verifizieren, dass sich Studierende im Forum stärker beteiligen als sie es in Präsenzsitzungen tun würden:

**Ich habe mich in den Foren durch das Schreiben von Beiträgen intensiver zu Wort gemeldet als in den Präsenzsitzungen. (n=28)**



Dieser Eindruck ist sowohl bei den ES als auch bei den DS erkennbar. Die erschwerte Kommunikation bei den virtuellen Präsenzsitzungen konnte darüber hinaus durch die Forenarbeit entlastet werden, was in folgender Aussage aus der Vorstudie deutlich wird:

*Der Austausch in den Seminaren empfand ich schwieriger, da man nicht sehen konnte, wer auch etwas beitragen möchte und das ist bei einer Präsenzsitzung einfach anders. Es war manchmal schwierig zu Wort zu kommen. Aber dennoch wurde das Semester von Ihnen gut gemeistert und man hat viel mitgenommen. Besonders die Forenbeiträge fand ich sehr bereichernd; ich hätte sonst nicht so viel dazu beitragen können.*

Allerdings müssen klarere Regularitäten bezüglich der Beiträge vorab festgelegt werden. In den freien Kommentaren werden u.a. folgende Aspekte erwähnt:

- Mehr Zeit für Kommentare
- Vorgaben über Länge der Beiträge z.B. durch Zeilenbegrenzung
- Redundanz vermeiden
- Moderation durch die Lehrenden.

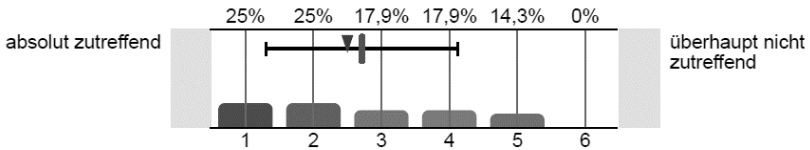
Als weiteres Zwischenfazit lässt sich festhalten, dass durch den Moduswechsel von der Mündlichkeit zur Schriftlichkeit, den digitale Lehre insbesondere durch die *ICM* bietet, sich eine vielversprechende Möglichkeit ergibt, auch weniger kommunikativen Lernenden eine Chance zur Beteiligung zu geben und dadurch eine Erhöhung des Lerneffekts zu ermöglichen.

Als letzter, jedoch nicht unerheblicher Bereich erfolgen im nächsten Kapitel die Ausführungen bezüglich der Maßnahmen, die eine soziale Isolation, auch digitale Distanz genannt, ausgleichen sollten.

## 2.5 Maßnahmen gegen die sog. digitale Distanz

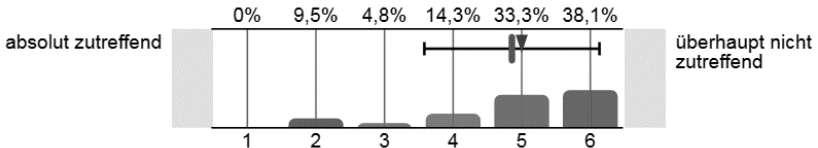
Verschiedene Maßnahmen bzw. Aktivitäten und Funktionen von Tools wurden eingesetzt, um die soziale Isolation der Studierenden durch die fehlenden realen Treffen zumindest abzuschwächen. Diese waren u.a. die Bedeutung der eingeschalteten Kamera während der Präsenzsitzungen (sowohl in Modell a als auch in b), die Integration der Chatfunktionen über *MS Teams* (bei Modell b) und der dadurch vermehrten Erreichbarkeit der Lehrenden. Überdies gibt die Häufigkeit der Teilnahme an Präsenzsitzungen Aufschluss darüber, als wie wertvoll diese angesehen werden. Dies wird durch den Fakt, dass die Anwesenheitspflicht ausgesetzt wurde, noch verstärkt. Beginnend bei Letzterem lässt sich festhalten, dass, obwohl in der *ICM* während der Videobesprechungen mit *MS Teams* keine relevanten Inhalte vermittelt wurden, die Anzahl der Teilnehmenden bis zum Ende des Semesters durchschnittlich bei ca. 75% lag. Außerdem wurden diese Besprechungen als weniger ermüdend empfunden als die 90 Minuten virtuelle Präsenzseminar aus Modell b:

**Ich empfand die regelmäßigen Sitzungen mit Zoom als zu ermüdend und habe nach kurzer Zeit mental abgeschaltet. (n=28)<sup>11</sup>**



Der Effekt ist bei den DS, die vermutlich noch mehr Videobesprechungen erlebt haben, mit einem Mittelwert von 2,9 zu 2,5 bei den ES noch ein wenig stärker. Dies deckt sich auch mit der allgemeinen „Zoom Fatigue“, die vor allem dann auftritt, wenn nur über Fachliches gesprochen wird und man sich zur Teilnahme an diesen Sitzungen verpflichtet fühlt (vgl. den Zeitungsartikel von Peter Ilg vom 16.09.2020). Für das Seminar zu Interkulturalität, bei dem die virtuellen Sitzungen „nur“ zum Austausch und häufig auch zum Plaudern genutzt wurden, zeigt sich dadurch ein deutlich anderes Bild:

**Ich empfand die Besprechungen in Teams als zu ermüdend und habe nach kurzer Zeit mental abgeschaltet. (n=21)**

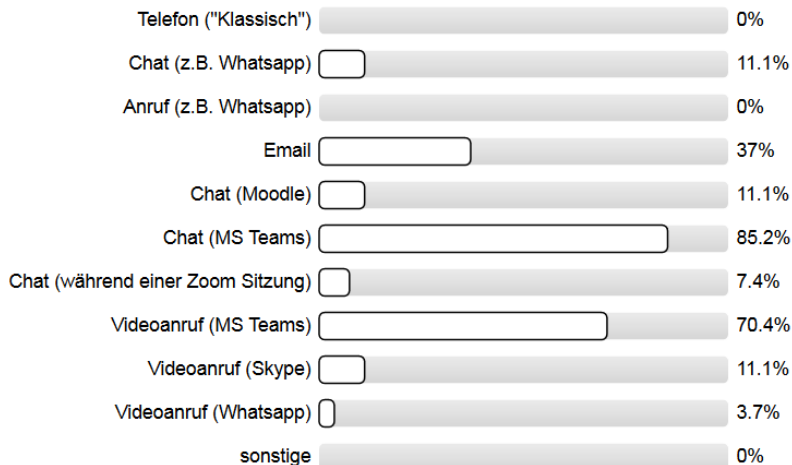


Dies spricht für das Modell b, in dem Kontakt zu gleichgesinnten KommilitonInnen hergestellt und durch den ungezwungenen Kontext eine soziale Distanz abgebaut werden konnte. Generell wird die Kontaktaufnahme über *MS Teams* als sehr gewinnbringend bewertet. Die Studierenden geben an, sich neben eigenen *Whatsapp* Gruppen (66,7%) über den Chat (48,1%) und den Vi-

.....

11 Auf die Negativität dieser Aussage wurde in der Umfrage explizit hingewiesen, weshalb davon ausgegangen werden kann, dass die Bewertungen bewusst gemacht wurden und damit aussagekräftig sind.

deanruf (40,7%) über *Teams* mit Ihren KommilitonInnen auszutauschen. Bei der Kontaktaufnahme zu den Dozentinnen überwiegt dagegen die Vorliebe für den Chat via *MS Teams* gegenüber der eher üblichen Form per Email (n= 27):



Die Erreichbarkeit, die durch die Nutzung von *MS Teams* vereinfacht wird, wirkte sich ebenso positiv auf die Bewertung der Studierenden aus, was vor allem in Kommentaren zum Ausdruck kommt:

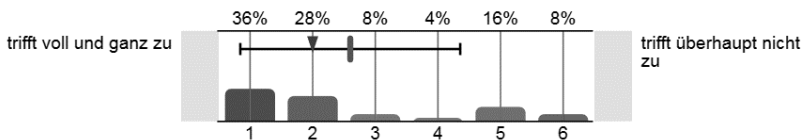
- *Die Erreichbarkeit von [...] war hervorragend. Fragen konnten viel unkomplizierter geklärt und Sprechstundentermine schneller organisiert werden. Gerade weil im 4. DaF-Semester ja keine bzw. kaum Präsenzveranstaltungen stattfinden und man weniger Zeit an der Uni verbringt, eignet sich MS Teams super, um in Kontakt mit der Uni, DozentInnen und Studierenden zu bleiben. Ich würde mir wünschen, dass diese Möglichkeit auch über die Coronazeit hinaus bestehen bleibt. Vor allem im Hinblick auf die Masterarbeit wäre es toll, die DozentInnen weiterhin so unkompliziert zu erreichen und bei Fragen ansprechen zu können.*



- *Die vielen Möglichkeiten der digitalen Lehre und der Kontaktaufnahme sind teilweise eine Bereicherung, aber ersetzen aus meiner Sicht nicht ganz den persönlichen, auch informellen Kontakt.*

Zusammenfassend kann festgehalten werden, dass sich gerade mit *MS Teams* auf besonders einfache Weise ein virtueller Campus simulieren lässt und dadurch helfen kann, die digitale Distanz abzuschwächen:

Durch die verschiedenen Funktionen von *MS Teams* wie Chatgruppen (z.B. Sprechstunden-Chat), die digitale Bibliothek, die Teams zu den einzelnen Seminaren usw. hatte ich das Gefühl, dass der Mangel an persönlichem Kontakt ausgeglichen werden konnte (n=25)



Doch auch hierbei wird deutlich, dass die Erfahrungen nicht bei allen Studierenden gleich waren und deshalb die Vielfalt der verschiedenen Kontaktmöglichkeiten angeboten werden muss. Auch der zeitliche Aufwand für die Vor- und Nachbereitung muss bei der Festlegung des Lehrhabitats berücksichtigt werden. Ein Vorteil, den der Chat dagegen bietet, ist die vereinfachte Kommunikation „auf Augenhöhe“, da sie eher der *face-to-face* Kommunikation ähnelt bzw. sich durch die sog. „Konzeptionelle Mündlichkeit“ (Albert 2013: 166ff.) auszeichnet. Dies kann dabei helfen, die angemessene Kommunikationsform mit (Hochschul-)Lehrenden z.B. in Bezug auf E-Mails zu finden, die sowohl für ausländische als auch für inländische Studierenden immer noch eine Herausforderung darstellt (vgl. Hiller 2014).

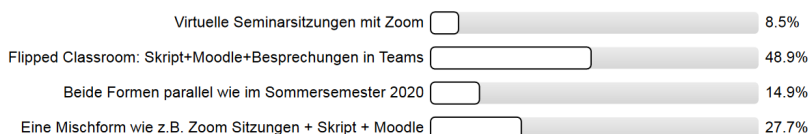
## 2.6 Zusammenfassender Vergleich der beiden Modelle

Ein valider Vergleich der beiden Modelle ist anhand der erhobenen Daten nicht möglich; dazu bedarf es einer detaillierteren Evaluation aller Seminare.

Außerdem muss ebenfalls die Diversität der Lehrformen bzw. der Lehrpersönlichkeiten beachtet werden. Nicht nur Studierende, sondern auch Lehrende haben ihre Vorlieben und ihren individuellen Lehrstil.

Dennoch lässt sich aufgrund der genutzten Medien und Tools eine Tendenz erkennen. Durch die sehr positive Bewertung der Videos, der Forenarbeit auf der Lernplattform, der Nutzung von *MS Teams* und den dortigen Chat-Funktionen und der synchronen Erreichbarkeit wird klar, dass Modell b zumindest für den Großteil der Studierenden für die rein digitale Lehre praktikabler ist. In der Online-Umfrage wurde gezielt nach der Präferenz der Studierenden für das Wintersemester gefragt:

**Wenn Sie wählen könnten, für welche Form würden Sie sich entscheiden?  
(n=46)**



Knapp die Hälfte der Studierenden wünscht sich das Modell b (hier als *Flipped Classroom* bezeichnet), nur gut 9% möchten ihre Seminare virtuell über *Zoom*, also Modell a, besuchen. Bezüglich dieser beiden Varianten gibt es bei den DS und den ES keine nennenswerten Unterschiede, lediglich wünschen sich mehr DS (n= 19) mit 42,1 % eine Mischform aus beiden Modellen.

Bezieht man Kommentare aus der Vorstudie und aus der Online-Umfrage in die Ergebnisse der Frage nach der Aufmerksamkeitsspanne (Kapitel 2.6) mit ein, ist auch hier eine deutlich positive Bewertung der nichtfachlichen, eher lockeren Besprechungen auf *Teams* zu erkennen. Sicher ist der Wunsch nach herkömmlicher Präsenzlehre bei allen Studierenden (und bei Lehrenden) gleichermaßen stark, aber Modell b scheint die Chancen, die digitale Lehre bietet, zu nutzen. Ferner werden 90 Minuten Seminare, wie sie ja in vielen Fachbereichen üblich sind, kritischer gesehen, d.h. die Studierenden reflektieren mehr über ihre eigenen Lehr- und Lernvorlieben, wie folgender Kommentar zeigt:

*Mit der Herangehensweise dieses digitalen Semesters habe ich festgestellt, dass es bei mir viel effektiver wirkt, wenn ich den Skript, Videos, Artikel und Forumbeiträge meiner Kommilitonen lesen und kommentieren muss als 90 Minuten im Unterricht sitzen und zuhören.*

Es zeigt wiederum, wie zentral die Individualisierung von Lehr-Lernprozessen in zeitgemäßer (nicht nur digitaler) Lehre ist und welche Bedeutung der soziale Austausch dabei einnimmt. Bei der didaktischen Umsetzung ist das *ICM* ein wichtiger Schritt: „The Inverted Classroom Model is not only the backbone to modern teaching and learning formats on an academic level but also solves several problems of traditional teaching and learning“ (Handke & Weber 2018: 131). Dies muss aber noch genauer und längerfristiger eruiert werden.

### 3 Fazit und Ausblick

Ausgangspunkt für diesen Beitrag war die Frage, auf welche Weise digitale Lehre unter erschwerten Bedingungen dennoch lernzielfördernd und motivierend eingesetzt werden kann. Eine spezifisch auf das Fach adaptierte Form des *ICM* bietet dafür einen vielversprechenden Ansatz, mit dem im Masterstudiengang DaF/DaZ sehr positive Erfahrungen gemacht werden konnten. In Bezug auf die eingangs erwähnte Frage nach Mitteln, Prioritäten, Rolle der Lehrenden und Voraussetzungen kann folgendes festgestellt werden:

- Skripte in der Kombination mit verschiedenen Aufgabenformen sind motivierend, praktikabel und können die Vermittlung der Lerninhalte, die sonst während Präsenzsitzungen erfolgt, ersetzen. Es muss dabei aber zwischen verschiedenen Aufgabentypen und Formen abgewechselt werden, ebenso müssen Lernziele wie z.B. bei Zusammenfassungen deutlich formuliert werden.
- Der Einsatz von Erklärvideos, sowohl rezeptiv als auch produktiv, trägt zur Motivation und zur Kompetenzentwicklung bei.
- Forenarbeit aktiviert alle Lernenden, sogar diejenigen, die in Präsenzseminaren eher zurückhaltender sind.

- *Microsoft Teams* bietet vielfältige Möglichkeiten zu Kommunikation in Echtzeit, die Simulation eines virtuellen Campus wird dadurch gegeben.
- Synchroner Termine, die nicht für Seminarinhalte, sondern zum Austausch genutzt werden, wirken der „Zoom-Fatigue“ entgegen und verringern das Gefühl der sozialen Isolation.
- Eine vermehrte Erreichbarkeit der Lehrenden z.B. durch Chats bewirkt ebenfalls, dass sich die Studierenden weniger allein gelassen fühlen. Die Kommunikation auf Augenhöhe verstärkt dies noch.

Im weiteren Verlauf gilt es dennoch, die Erkenntnisse gerade bezüglich der digitalen Distanz noch weiter zu vertiefen. Dies betrifft vor allem:

- die Erweiterung der Erklärvideos mit barrierefreiem Zugang z.B. über einen offenen *Youtube*-Kanal.
- die Implementierung weiterer Interaktions-Tools wie *gather.town* zum Aufbau des „sozialen Dorfs“.
- Erarbeitung einer erweiterten Feedback-Kultur, die für Lehrende zeitlich und organisatorisch händelbar ist und nicht auf Dauer zu einer Überbelastung und „ständigen“ Erreichbarkeit führt.

Letztere ist aber für die Lehrenden ein erheblicher Mehraufwand, da diese intensive Betreuung, neben der Konzeption des Online-Unterrichts, sehr zeitaufwändig ist. Handke & Weber (2018: 138) machen das auch noch einmal deutlich: „Die Rolle des Lernbegleiters im Hörsaal – seine Motivation, seine Kenntnisse und sein Geschick – sind nach unserer Einschätzung nach wie vor gefragt, in einem Inverted-Classroom-Szenario übrigens mehr denn je zuvor.“ Dies muss sich auch auf die finanziellen Lage auswirken: „Die Bundesländer, welche Regelungen zu digitaler Lehre berücksichtigen, sehen eine Anrechnung der Durchführung von EL-Angeboten und deren Betreuung auf das Lehrdeputat vor“ (Lungershausen, Emunds & Buß 2016: 102). Dies wird bis dato aber scheinbar nicht umgesetzt, obwohl Beispiele für entsprechende Modelle der Deputatsanrechnungen für Formen des digitalen Lernens bereits vorliegen

(vgl. ebd.). Eine Befragung der Lehrenden wäre also ebenfalls sinnvoll<sup>12</sup>, um mögliche Neuberechnungen zu fundieren und damit auch noch mehr Anreize zu schaffen, um weniger über die Machbarkeit von Präsenzterminen und den damit einhergehenden Hygienevorschriften zu diskutieren, sondern die aktuellen Chancen mehr dazu zu nutzen, sich auf neue Lernformate einzulassen.

## Bibliografie

- ALBERT, G. (2013): *Innovative Schriftlichkeit in digitalen Texten. Syntaktische Variation und stilistische Differenzierung in Chat und Forum*, Berlin: Akademie Verlag, abrufbar unter: <https://doi.org/10.1524/9783050063676>
- BÜHNER, B. & SOMMER, J. (2020): „Der Inverted Classroom – eine Königsdisziplin der digitalen Hochschullehre?“, in: FRIEDERICHSEN, M. & WERSIG, W. (Hrsg.): *Digitale Kompetenz*, Wiesbaden: Springer Fachmedien: 129–134.
- CHEN, C.-M., WANG, J.-Y. & YU, C.-M. (2017): „Assessing the Attention Levels of Students by Using a Novel Attention Aware System Based on Brainwave Signals“, in: *British Journal of Educational Technology* 48: 348–369.
- DIGITALPAKT SCHULE: *Verwaltungsvereinbarung DigitalPakt Schule 2019 bis 2024* (o.S.), abrufbar unter: [https://www.digitalpaktschule.de/files/VV\\_DigitalPaktSchule\\_Web.pdf](https://www.digitalpaktschule.de/files/VV_DigitalPaktSchule_Web.pdf), 23.09.2020.
- ERNST, D. (2017): „Möglichkeiten eines gesprächsanalytischen Zugangs zu Erklärkompetenz“, in: HAUSER, L. (Hrsg.): *Gesprächskompetenz in schulischer Interaktion – normative Ansprüche und kommunikative Praktiken*, Mündlichkeit, Band 5, Bern: hep verlag ag: 158–188, abrufbar unter: <https://www.zg.ch/behoerden/direktion-fur-bildung-und-kultur/phzg/forschung/zm/publikationen/downloads/gespraechskompetenz-in-schulischer-interaktion/@@download/file/Gespraechskompetenz%20in%20schulischer%20Interaktion.pdf#page=157>, 03.10.2020.

.....

12 Einen ersten Schritt dahingehend unternimmt Grein in diesem Band.

- GREIN, M. & STRASSER, T. (2019): „Lernen mit digitalen Medien aus neurobiologischer und fremdsprachendidaktischer Sicht“, in: *Empfehlungen Goethe-Institut Zagreb*: 7–15.
- GREIN, M. (2020): „Sprachunterricht auf einmal digital: Online-Lernen vor, während und nach Corona“, in: *Magazin Sprache*, Goethe Institut: o.S., abrufbar unter: <https://www.goethe.de/de/spr/mag/21927962.html>, 09.09.2020.
- HANDKE, J. & WEBER, K. (2018): „Lernerverhalten im Inverted Classroom. Eine Lehrveranstaltung auf dem Prüfstand“, in: BUCHNER, J. et al. (Hrsg.): *Inverted Classroom. Vielfältiges Lernen. Begleitband zur 7. Konferenz Inverted Classroom and Beyond* (2018), Brunn am Gebirge: Ikon Verlag: 131–139.
- HANDKE, J. (2020): *Handbuch Hochschullehre Digital: Leitfaden für eine moderne und mediengerechte Lehre*, 3., aktualisierte und erweiterte Auflage, Baden-Baden: Tectum Verlag.
- HILLER, G. G. (2014): „Kulturelle und sprachliche Diversität in der Hochschule am Beispiel von E-Mail-Kommunikation“, in: MOOSMÜLLER, A. & MÖLLER-KIERO, J. (Hrsg.): *Interkulturalität und kulturelle Diversität*, Münster: Waxmann: 233–258.
- ILG, P. (2020): „Zoom-Fatigue: Virtuelle Meetings machen müde“, in: HEISE ONLINE: o.S., abrufbar unter: [https://www.heise.de/news/Zoom-Fatigue-Virtuelle-Meetings-machen-muede-4892703.html?utm\\_source=pocket-newtab-global-de-DE](https://www.heise.de/news/Zoom-Fatigue-Virtuelle-Meetings-machen-muede-4892703.html?utm_source=pocket-newtab-global-de-DE), 03.10.2020.
- KERGEL, D. & HEIDKAMP-KERGEL, B. (2020): *E-Learning, E-Didaktik und digitales Lernen*, Wiesbaden: Springer VS, abrufbar unter: <https://doi.org/10.1007/978-3-658-28277-6>, 03.10.2020.
- LEENEN, W.-R. (2019): „Grundbegriffe interkultureller Kompetenzvermittlung“, in: DERS. (Hg.): *Handbuch Methoden interkultureller Weiterbildung*, Göttingen: V&R: 25–167.
- LUNGERSHAUSEN, U., EMUNDS, G. & BUSS, I. (2016): „Anrechnung virtueller Lehre auf das Lehrdeputat“, in: *Die neue Hochschule 4*: 102–105, abrufbar unter: [http://h1b.de/fileadmin/h1b-global/downloads/dnh/full/2016/DNH\\_2016-4.pdf#page=8](http://h1b.de/fileadmin/h1b-global/downloads/dnh/full/2016/DNH_2016-4.pdf#page=8), 21.10.2020.

- MARX, N. & LANGER, M. (2005): „Einleitung zum Themenheft: Die neuen Medien im Fremdsprachenunterricht“, in: *ZIF 10/1*: o.S., abrufbar unter: <https://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/407/395>, 30.07.2020.
- MAXIN, J.H. (2018): „Videografie in Schule und Unterricht am Beispiel einer Interventionsstudie zur Erklärkompetenz von RealschülerInnen“, in: MORITZ, C. & CORSTEN, M. (Hrsg.): *Handbuch Qualitative Videoanalyse*, Wiesbaden: Springer Fachmedien: 607–622, abrufbar unter: [https://doi.org/10.1007/978-3-658-15894-1\\_32](https://doi.org/10.1007/978-3-658-15894-1_32), 03.10.2020.
- MEISSELBACH, C. & BOCHMANN, C. (2020): *Studierendenbefragung zur digitalen Lehre in Zeiten der Corona-Krise*, Dresden, abrufbar unter: [https://tu-dresden.de/gsw/phil/powi/polsys/ressourcen/dateien/forschung/umfrage-digitale-lehre/studierendenbefragung-digitale-lehre.pdf/at\\_download/file](https://tu-dresden.de/gsw/phil/powi/polsys/ressourcen/dateien/forschung/umfrage-digitale-lehre/studierendenbefragung-digitale-lehre.pdf/at_download/file), 24.09.2020.
- MERTENS, C. et al. (2019): „To flip or not to flip?“ Empirische Ergebnisse zu den Vor- und Nachteilen des Einsatzes von Inverted-Classroom-Konzepten in der Lehre“, in: SCHMOHL, T. & TO, K. A. (Hrsg.): *Hochschullehre als reflektierte Praxis. Fachdidaktische Fallbeispiele mit Transferpotenzial*, Bielefeld: wbv: 13–28.
- MODULHANDBUCH für den Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache, Johannes Gutenberg-Universität Mainz, 3/2019, abrufbar unter: <https://www.daf.uni-mainz.de/Texte/Modulhandbuch.pdf>, 03.10.2020.
- ORTMANN-WELP, E. (2020): „Die Potenziale digitaler Medien für Lernprozesse“, in: DERS.: *Digitale Lernangebote in der Pflege. Neue Wege der Mediennutzung in der Aus-, Fort- und Weiterbildung*, Berlin/Heidelberg: Springer: 13–22.
- SCHEITER, K. & LACHNER, A. (2019): „DigitalPakt – was nun? Eine Positionierung aus Sicht der Lehr-Lernforschung“, in: *Unterrichtswiss 47*: 547–564, abrufbar unter: <https://doi.org/10.1007/s42010-019-00059-2>, 09.09.2020.
- SCHMIDT-BORCHERDING F. et al. (2020): „Erklärvideos im inverted classroom – Forschungsmethoden vermitteln im Masterstudiengang Erziehungs- und Bildungswissenschaften“, in: HOFFMEISTER T. et al. (Hrsg.):

*Forschendes Lernen als Studiengangprofil*, Wiesbaden: Springer VS, abrufbar unter: [https://doi.org/10.1007/978-3-658-28825-9\\_17](https://doi.org/10.1007/978-3-658-28825-9_17), 03.10.2020.

TRAUS, A. et al. (2020): *Stu.di.Co. – Studieren digital in Zeiten von Corona. Erste Ergebnisse der bundesweiten Studie Stu.di.Co.*, Hildesheim: Universitätsverlag, abrufbar unter: [https://hildok.bsz-bw.de/files/1126/Thomas\\_Stu.di.Co.pdf](https://hildok.bsz-bw.de/files/1126/Thomas_Stu.di.Co.pdf), 23.09.2020.

## Internetquellen

CHANCEN JGU (2020): *Chancen und Herausforderungen digitaler Lehre* (Erfahrungsberichte aus der JGU), abrufbar unter: <https://lehre.uni-mainz.de/digital/chancen/>, 24.09.2020.

RUBEL (o.J.): *Inverted Classroom* (eLearning (RUBeL)), Zentrum für Wissenschaftsdidaktik, Ruhr-Universität Bochum, abrufbar unter: <https://dbs-lin.ruhr-uni-bochum.de/lehreladen/e-learning-technik-in-der-lehre/inverted-classroom/was-ist-inverted-classroom/>, 04.10.2020.





MARION GREIN

## Evaluation der unterschiedlichen virtuellen Sprachkursformen

### Einleitung

Der Masterstudiengang Deutsch als Fremdsprache/Deutsch als Zweitsprache der Johannes Gutenberg-Universität Mainz ist einer der wenigen Studiengänge, bei dem Lernende bzw. zukünftige Lehrende ein Modul zur „Digitalen Lehre“ absolvieren. Dies mag der Grund für zahlreiche Anfragen zur Beratung im Bereich „Digitale Lehre“ zu Beginn der Corona-Krise gewesen sein. In diesem Zusammenhang gab es sowohl Anfragen der Volkshochschulen, des Goethe-Instituts und auch verschiedener weiterer Organisationen (u.a. bei der Tagung Fremdsprachenlernen im digitalen Zeitalter/Auswirkungen von Covid-19 auf das (Fremdsprachen)Lernen des Goethe-Institut Peking und dem Institut français im Rahmen des Deutsch-Französischen Kulturfonds (Elysee-Fond), den „Jornadas Académicas“ in Pachuca, Mexiko, dem Fremdsprachenzentrum der Universität Passau und der GIP-Tagung der Pädagogischen Hochschule Freiburg: „Lehr-Lern-Medien für DaF aus nationaler und internationaler Perspektive“. Diese Panel-Diskussionen, Vorträge und Fortbildungen ließen erkennen, dass die Umsetzung der digitalen Lehre sehr unterschiedlich realisiert worden ist und waren Anlass für eine erste kleine Studie zur Akzeptanz der insgesamt fünf ausgemachten Formen des virtuellen Sprachunterrichts. Im Zentrum meiner Vorträge stand die Notwendigkeit des Ausgleichs der sog. digitalen Distanz, die lediglich in einem Kursformat im Vordergrund steht. Überprüft wurde mit der kleinen Studie, inwieweit der Ausgleich der Distanz durch zahlreiche soziale und kollaborative Tools tatsächlich erfolgreicher ist als andere Formen des virtuellen Sprachunterrichts.

In einem ersten Kapitel wird knapp auf die Ausgangssituation vor Corona eingegangen. Es folgt im 2. Kapitel eine Thematisierung der sog. digitalen Distanz, die es auszugleichen gilt (vgl. Studie Ersch 2020). Im 3. Kapitel werden die fünf unterschiedlichen Typen des virtuellen Unterrichts skizziert. Dabei wird auch ein kurzer Blick auf die genutzten Video-Chat Tools geworfen. Das 4. Kapitel präsentiert die Ergebnisse der sehr kurzen Prästudie. Diese dient vor allem auch den Studierenden des Masterstudiengangs DaF/DaZ, die in diesem Bereich ihre Masterarbeiten verfassen. Hier werden spezifischere empirische Studien angefertigt, deren Ergebnisse dann einen zusätzlichen Erkenntnisgewinn versprechen.

## 1 Digitales Lernen vor Corona

Die Analyse von Persike & Friedrich (2016) zum Stellenwert der digitalen Lehre in Deutschland im Auftrag des Hochschulforums Digitalisierung macht deutlich, dass digitale Medien an der Mehrzahl der Hochschulen kein integrierter Bestandteil der Lehre sind. De Florio-Hansen (2018: 291) schrieb mit Bezug zum Fremdsprachenunterricht, dass gerade dieser Bereich sehr von der Digitalisierung profitieren kann. Wenn jedoch gerade an den Hochschulen der Umgang mit digitalen Medien als sehr gering eingestuft wird und darüber hinaus auch kein Bestandteil der Lehrendenausbildung ist, wie soll er dann an die Schulen oder Sprachschulen kommen? Wie eingangs dargestellt, ist das Modul zur „Digitalen Lehre“ im Masterstudium DaF/DaZ an der Universität Mainz fast ein Alleinstellungsmerkmal. Van Ackeren et al. (2019: 110) fassen den Stand der Digitalisierung zusammen:

In seiner Zusammenschau kommt der Monitor Lehrerbildung in seiner 2018 vorgelegten Studie zu dem Schluss, dass insbesondere das Lehramtsstudium in den Bundesländern derzeit weit davon entfernt sei, Inhalte des Studiums zum Umgang mit digitalen Medien im Sinne der von der Kultusministerkonferenz angestrebten Ziele verbindlich festzuschreiben.

Aktuell fordern Lorenz & Endberg (2019: 62) die gezielte Nutzung digitaler Medien für das fachliche und überfachliche Lernen. Sie argumentieren ebenfalls, dass die Nichtnutzung digitaler Tools in „hohem Masse an der fehlenden professionellen Handlungskompetenz der Lehrpersonen im Kontext der Digitalisierung“ liegt. Dies wiederum liegt an den fehlenden Möglichkeiten, diese Kompetenzen in der Lehramtsausbildung zu erwerben (vgl. Eickelmann et al. 2016; Schiefner-Rohrs 2018).

Für den Bereich Sprachen bot gerade das Goethe-Institut zahlreiche Weiterbildung für Lehrende an (z.B. das u.a. aufgeführte *DIP in SOE – digital, interkulturell – projektorientiert in Süd-Ost-Europa*). Auch auf der Plattform *Dhoch3* des DAAD findet sich ein Modul zum digitalen Lernen (Modul 3: Lehren und Lernen mit elektronischen (Online-)Medien).

Wie Grein & Strasser (2019) im Rahmen des Projekts *DIP in SOE* thematisieren, wurde im Bereich Sprachunterricht der sinnvolle Einsatz unterschiedlicher digitaler Tools auch im *face-to-face*-Unterricht hervorgehoben. Aufgegriffen wurden dort (2019: 17f.) verschiedene *Educational Applications (EduApps)*, also eigens „für unterrichtliche Zwecke entwickelte browser-basierende Internettools“ (Schmidt & Strasser 2016: 5). Aus fremdsprachendidaktischer Sicht basieren *EduApps* auf folgenden Prinzipien (vgl. Schmidt & Strasser 2016, zitiert in Grein & Strasser 2019):

1. Reflexion: Brainstorming- bzw. Mindmapping-Anwendungen wie z.B. *EduPad*<sup>1</sup> oder *Padlet*<sup>2</sup> unterstützen die Lernenden bzw. Lehrenden in der Reflexion bestimmter Arbeitsprozesse (z.B. Feedback zur sprachlichen Performanz beim gemeinsamen Verfassen einer Abenteuergeschichte).
2. Modifikation: Werkzeuge wie *Quizlet*<sup>3</sup> oder *Kahoot*<sup>4</sup> ermöglichen es digitale Lernprodukte (z.B. Flashcards) zu produzieren, um diese dann an ihre Lehr- und Lernbedürfnisse anzupassen.

.....

1 [www.edupad.ch](http://www.edupad.ch)

2 [www.padlet.com](http://www.padlet.com)

3 [www.quizlet.com](http://www.quizlet.com)

4 [www.create.kahoot.com](http://www.create.kahoot.com)

3. Kommunikation: Mit Anwendungen wie *Edupad*<sup>5</sup>, *Today'smeet*<sup>6</sup>, *Tricider*<sup>7</sup> etc. findet Kommunikation über den Lernprozess zwischen den Lernenden statt. Heute bietet sich hier vor allem *Gathertown*<sup>8</sup> als sinnvolles Tool zur Kommunikation an.
4. Multiplikation: Die Lernprodukte (z.B. Collage mit *Padlet*) können auf einfache Art und Weise mit Mitlernenden und der Lehrkraft geteilt werden.
5. Kreation: Lernende und Lehrende können einfache Lern- und Unterrichtsmaterialien selbst erstellen (z.B. Collagen mit *Padlet*, Vokabellernkarteien mit *Quizlet*). Lernende sind aktive ProduzentInnen des Lernprozesses.
6. Kollaboration: Lehrende und Lernende arbeiten aktiv innerhalb eines digitalen Frameworks zusammen, um lektionsspezifische und ergänzende Aufgaben des möglichst digital vorliegenden Lehrwerks zu erledigen.

Schon in den Jahren zuvor wurde deutlich, dass gerade Blended-Learning Konzepte dann zu mehr Lernerfolg führen als reiner *face-to-face*-Unterricht, wenn die Lernenden zu kooperativem Lernen aufgefordert wurden, sich als Gruppe identifizierten und sie aktiv durch die Lehrkraft unterstützt wurden (vgl. Appel & Schreiner 2014: 5). Als besonders erfolgreich schnitten dabei auch bereits 2014 sog. *Serious Games*, bei denen Lernende selbst durch das Spiel navigieren können, ab (vgl. ebd.), immer unter der Prämisse, dass Lernende auch zum kollaborativen Austausch angeregt wurden und die Lehrkraft als Ansprechpartner/in zur Verfügung stand. Auch im Bereich DaF/DaZ wurden die Forderungen für den Bereich der Digitalisierung aufgegriffen: Funk (2019: 76) differenziert zwischen konsumtivem, reaktivem, konstruktivem und konstruktiv-kollaborativem Einsatz der digitalen Medien. Konsumieren bedeutet dabei beispielsweise Filme und Erklärvideos betrachten, reaktiv dann die

.....

5 [www.edupad.ch](http://www.edupad.ch)

6 [www.todaysmeet.com](http://www.todaysmeet.com)

7 [www.tricider.com](http://www.tricider.com)

8 <https://gather.town/>

zahlreichen auch in Lehrwerke integrierten Übungsformen (z.B. Zuordnung, Multiple-Choice), die ein direktes (einfaches) Feedback geben. Konstruktiv und konstruktiv-kollaborativ ist dann das Arbeiten der Lernenden mit den Tools oder *EduApps*, also der konstruktivistische Einsatz der digitalen Medien. Die gleiche Forderung finden wir bei Roche (2019: 92f.), wenn er formuliert: „Die heutige Technologie müsste den Lernern nicht nur Erlauben, Sprache zu rezipieren oder zu imitieren, sondern sie auch befähigen, Sprache kreativ für Ausbildung und Alltag zu produzieren“. Und auch Rösler (2019: 14 FN) fordert zum aktiven Umgang mit den digitalen Tools auf und argumentiert: „Mit der Aufgabenorientierung sollen die Lernenden dazu gebracht werden, motiviert eigene Aktivitäten in der fremden Sprache vorzunehmen“.

Aus diesen Erkenntnissen entwickelten wir an der Johannes Gutenberg-Universität das in Ersch (dieser Band) dargestellte Konzept des *Inverted Classroom* für unsere Studierenden und evaluierten dieses. Besonders positiv bewertet wurde dort die Möglichkeit des sozialen Austauschs, das kollaborative Zusammenarbeiten und die Erreichbarkeit der Dozierenden. Aus diesen Erfahrungen und den o.a. *EduApps* wurde ein Konzept auch für den Fremdsprachenunterricht entwickelt und in den angesprochenen Weiterbildungen vorgestellt und ebenso veröffentlicht (vgl. Grein 2020a, Grein 2020b). Das Konzept wird im nächsten Kapitel kurz skizziert.

## 2 Konzept zum Ausgleich der digitalen Distanz

In Grein (2020a: o.S.) wird deutlich formuliert, dass die soziale Eingebundenheit und der soziale Austausch zentrale Faktoren für den Lernerfolg sind. In Anlehnung an Alexiou-Ray & Bentley (2016) wird argumentiert, dass die sog. digitale Distanz durch einen erhöhten Kontakt unter den Teilnehmenden und den Lehrkräften ausgeglichen werden muss. Dabei geht es vor allem um das Gefühl, sich austauschen zu können, nicht unbedingt um die Frequenz der Nutzung des Angebots. Vorgeschlagen wird der Einsatz einer Lernplattform (LMS), bei der die soziale Aktion im Vordergrund steht und nicht die Kontrolle der Leistungen.

Um die Aufmerksamkeitsspanne der Teilnehmenden nicht außer Acht zu lassen, wird empfohlen, die digital vorliegenden und inzwischen editierbaren Lehrwerke über *Screensharing* zu nutzen. In Grein (2020b) findet sich ein Vorschlag für die zeitliche Einteilung eines 5-stündigen Intensivkurses (Abb. 1), den man dann an die zur Verfügung stehenden Zeit anpassen kann. Zentral sind dabei die Phasen, in denen die Teilnehmenden kollaborativ zusammenarbeiten. Dabei kann man sowohl die Übungen und Aufgaben aus dem Lehrwerk verwenden, aber auch die o.a. *EduApps*. Konkrete Beispiele folgen. Grundsätzlich wird empfohlen den Plan den Bedürfnissen der Lernenden anzupassen, also Flexibilität bei der Gestaltung des Online-Unterrichts zuzulassen. Je nach Lerntraditionen oder der vorhandenen Medienkompetenz bedarf es einer Anpassung vor allem der Zeiten für die Erklärungen der zum Einsatz kommenden *EduApps*.

Wie sieht der vorgeschlagene Ablauf aus?

10:00-10:10 Begrüßung, Austausch, Fragen klären	Plenum im virtuellen Raum
10:10-10:50 Lehrwerkbasierter Unterricht	Plenum im virtuellen Raum; Versuch alle zur Mitarbeit zu aktivieren
10:50-11:15 Partner- & Gruppenarbeiten (aus Lehrwerk)	Virtuelle Kleinräume
11:15-11:30 Präsentation der Ergebnisse	Plenum im virtuellen Raum
11:30-11:45 PAUSE	Anregung zu Bewegung Lehrkraft steht im virtuellen Raum zur Verfügung
11:45-12:15 Lehrwerkbasierter Unterricht	Plenum im virtuellen Raum; Versuch alle zur Mitarbeit zu aktivieren
12:15-12:30 Aufgabenstellung für Partner- Gruppenarbeiten (binnendifferenzierend)	Plenum im virtuellen Raum

12:30-13:15 Partner- und Gruppenarbeit zur Wiederholung z.B. mit digitalen Tools	Virtuelle Kleinräume
13:15 Treffen im Plenum; Diskussion wie viel Zeit die Gruppen noch benötigen	Plenum im virtuellen Raum
13:30-14:00 Mittagspause	
14:00-14:15 Lehrkraft wartet im virtuellen Raum, je nach weiterhin benötigter Zeit und „redet“ mit den Anwesenden	
14:15-15:00 Lernende präsentieren Ihre Zusammenfassungen des Lernstoffs; die anderen probieren die „Aktivitäten“ der jeweils anderen Gruppen	Plenum im virtuellen Raum
15:00 Zusammenfassung und Dank durch die Lehrkraft - Verabschiedung	

Abb. 1: Möglicher Aufbau eines 5-stündigen Online-Kurses

Die Aufgabenstellung muss präzise sein und möglichst unterschiedliche *Edu-Apps* sollten angeboten werden, daher benötigt die Aufgabenstellung tatsächlich oftmals 15 Minuten.

Wie kann das nun konkret aussehen? Auszug aus dem Lehrwerk *Momente* (Hueber):



10 | (10 von 109)

Originalgröße

2 **Wie heißt du?** AB 1-3

a Hören Sie noch einmal und ergänzen Sie.

Wie ich bin heißt wer?

- Wie heißt du?
- Ich bin Andrea.
- Ah, du bist Andrea.
- Andrea. A - N - D - R - E - A.
- Und        bist du?
- heiße Miriam.

heißen	sein
ich	heiße
du	heißt
	bist

Bitte arbeiten Sie alleine

b Das Alphabet. Hören Sie und sprechen Sie nach.

A, a [A]	B, b [Be]	C, c [Ce]	D, d [De]	E, e [E]	F, f [Fe]	G, g [Ge]	H, h [He]	I, i [I]	J, j [Je]
K, k [Ka]	L, l [Le]	M, m [Me]	N, n [Ne]	O, o [O]	P, p [Pe]	Q, q [Qe]	R, r [Re]	S, s [Se]	T, t [Te]
U, u [U]	V, v [Ve]	W, w [We]	X, x [Xe]	Y, y [Ye]	Z, z [Ze]	Ä, ä [Ä-Umlaut]	Ö, ö [Ö-Umlaut]	Ü, ü [Ü-Umlaut]	ß [Es-Zett]

Namen buchstabieren mit Konsonanten = Arme hoch; Vokale = Arme zur Seite

Abb. 2: Auszug aus dem Lehrwerk *Momente* (Hueber Verlag) mit integrierten Anweisungen/ Tipps (Original in Farbe)

Mit den Editiertools können Arbeitsanweisungen direkt in das Lehrwerk integriert werden. Aktivierende Übungen, wie hier das bewegte Buchstabieren, können als Tipp im Lehrwerk notiert werden, so dass auch im Plenum „Bewegung“ stattfindet.

Alle Hörtexte, Filme und reaktive Übungen können sofort gestartet werden. Bei den Hörtexten und Filmen kann man darüber hinaus die Geschwindigkeit regeln. Hier kann dann die erste *EduApp Gathertown*, bei der die Lernenden sich durch den Raum bewegen, zum Einsatz kommen.

c Hören Sie. Machen Sie dann einen Kurzspaziergang und fragen Sie drei Personen. Notieren Sie die Vornamen und die Familiennamen.

- Hallo. Ich heiße David Ramos.
- Entschuldigung. Wie bitte?
- David Ramos.
- D-a-v-i-d R-a-m-o-s.
- Danke.
- Und wer bist du?
- Mein Name ist ...

Vorname	Familiennamen
1 David	Ramos
2 ...	

Hier kann man bereits in gather.town wechseln [https://gather.town/app/i1bmpNthlmpB88hU/Marion\\_Grein](https://gather.town/app/i1bmpNthlmpB88hU/Marion_Grein)

3 **Wie geht's?** AB 4

a Wie geht es Andrea und Miriam? Hören Sie und kreuzen Sie an.

Super! / Sehr gut!
  Gut.
  Es geht.
  Nicht so gut.

Abb. 3: Auszug aus dem Lehrwerk *Momente* (Hueber Verlag) mit integrierten Anweisungen/ Tipps (Original in Farbe)

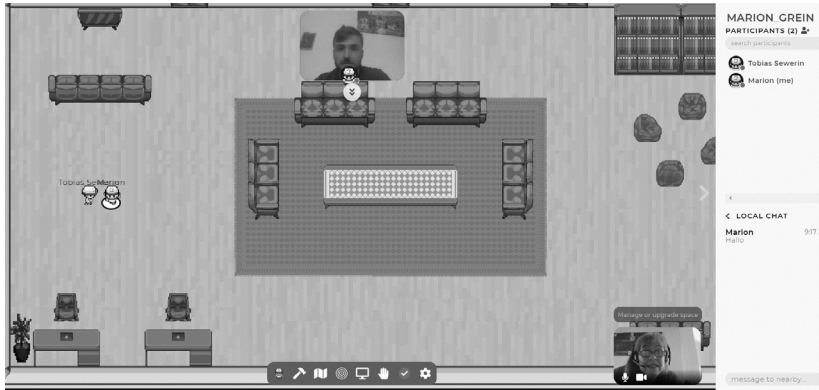


Abb. 4: Screenshot aus *Gathertown* (Original in Farbe)

In *Gathertown* werden Kamera und Mikrofon erst aktiv, wenn man nahe beieinandersteht, so dass man hier Partner- und Gruppenarbeiten durchführen kann. Die Teilnehmenden loggen sich ein und laufen mit den Pfeiltasten durch den Raum, um auf andere zu treffen.

Ein weiteres Tool sind *learning.apps*<sup>9</sup>, mit dem die Teilnehmenden unterschiedliche Übungsformen jeweils für die anderen erstellen können (Abb. 5). Hier wiederholen die Teilnehmenden und überlegen, welche der vielen Übungsformen sie dafür verwenden wollen.

.....  
9 <https://learningapps.org/>

**4 Das ist Simon.** AB 5-9

**a** Wer sagt was? Hören Sie das Gespräch weiter und ergänzen Sie:  
Andrea = A, Miriam = M, Simon = S

- 1  Woher **kommst du**, Miriam?
- 2  Ich komme aus Eritrea. Und du?
- 3  Ich komme aus Deutschland.
- 4  Wer ist das?
- 5  Das ist Simon. Das ist Miriam.
- 6  Sie kommt aus Eritrea. Sie lernt Deutsch.
- 7  Er heißt Simon. Er kommt aus der Schweiz.
- 8  Tschüs, Andrea.
- 9  Auf Wiedersehen, Andrea!

**b** Markieren Sie *ich, du, er/sie* und die Verben in a und ergänzen Sie dann.

	kommen	lernen	heißen	sein
ich	komm...	lerne	heiße	bin
du	kommst	lernst	heißt	bist
er/sie	komm...	lern...	heiß...	ist.

Simon → er  
Miriam → sie

<https://learningapps.org/watch?v=pt6c0adj320>

Abb. 5: Auszug aus dem Lehrwerk *Momente* (Hueber Verlag) mit integrierten Anweisungen/ Tipps (Original in Farbe)

In Form von „Wer wird Millionär“ (Abb. 6) können die Lernenden Verständnisfragen selbst „programmieren“ und mit den anderen teilen. Da die unterschiedlichen Gruppen andere Übungstypen wählen, hat man hier eine gute Form der Wiederholung, da die Lernenden auch die Übungen ihrer Mitlernenden lösen.

**Woher kommt Simon?**

**A** aus Österreich

**B** aus der Schweiz

**C** aus Deutschland

**D** aus Eritrea

Abb. 6: Übung zum Lehrwerktext erstellt mit *learning.apps* (Original in Farbe)



Für das Plenum kann man wieder „Bewegung“ in den Unterricht bringen:



**1 Ich heiße Lydia.**

**a** Sehen Sie das Bild an. Was sagen die Personen?  
Was meinen Sie? Kreuzen Sie an.

Arno und Lydia ...

1  sind 28 (= achtundzwanzig) Jahre alt.  
 sind 38 (= achtunddreißig) Jahre alt.

2  sind verheiratet.  

3  haben Kinder.   haben keine Kinder. 

4  wohnen in Berlin.  wohnen in München.  wohnen in Wien.

**b** Hören Sie und vergleichen Sie.

**LERNZIELE**

**KOMMUNIKATION:** über den Beruf und Persönliches sprechen: *Ich bin Friseurin ...* • den Wohnort nennen: *Sie wohnen in Berlin.* **HÖREN/LESEN:** Interview **SCHREIBEN:** ein Internetprofil ergänzen • einen kurzen Text über sich schreiben **WÖRTER:** Berufe, Familienstand, Zahlen: 1–100 **GRAMMATIK:** Konjugation Singular und Plural: *wohnen, leben, haben, sein ...* • Negation mit *nicht* • Präpositionen: *als, bei, in* • Wortbildung: Endung *-in*

Abb. 7: Auszug aus dem Lehrwerk *Momente* (Hueber Verlag) mit integrierten Anweisungen/ Tipps (Original in Farbe)

Für das kollaborative Arbeiten stehen zahlreiche Tools zur Verfügung:

**b** Schreiben Sie fünf Sätze über Lydia und Arno.

Die sind Lydia und Arno. Sie sind ...

**c** Würfelspiel: Wir leben in Berlin. Arbeiten Sie in Gruppe ...

**3 Wie alt seid ihr?** AB 4.4

**a** Hören Sie und sprechen Sie nach.  
Hören Sie dann noch einmal und notieren Sie die fehlenden Zahlen.

14: vierzehn

Grammatik Erklärvideo

<https://online-wuerfel.de/>

edupad <https://edupad.ch/p/VXK5IcM5PQ>

Lydia + Arno → sie

	wohnen	leben	sein	haben
sie	wohnen	leben	sind	haben

0	1	2	3	4	5	6	7	8	9	10	11	12
null	eins	zwei	drei	vier	fünf	sechs	sieben	acht	neun	zehn	elf	zwölf

Abb. 8: Auszug aus dem Lehrwerk *Momente* (Hueber Verlag) mit integrierten Anweisungen/ Tipps (Original in Farbe)

Zum einen kann man über das Internet würfeln. Kollaborativ und mit großer Kreativität (oder auch nicht → Binnendifferenzierung) schreiben die Teilnehmenden gemeinsam einen Text zu Lydia und Arno (siehe Abb. 8) auf [edupad.ch](https://edupad.ch/)<sup>10</sup>.

<sup>10</sup> <https://edupad.ch/> (Hier ein B2-Kurs zum Testen).

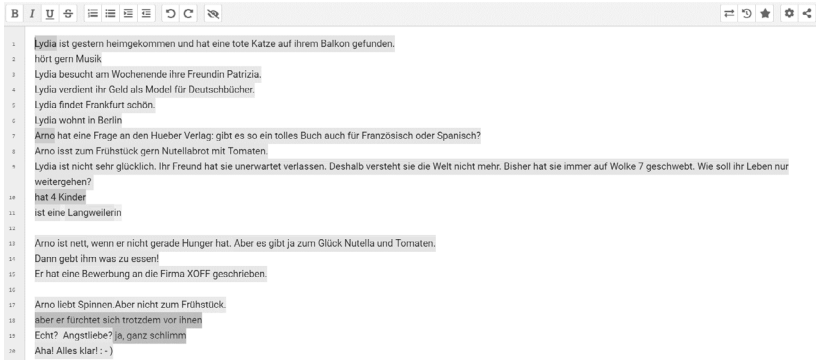


Abb. 9: *Edupad* zu zwei Protagonisten aus dem Lehrwerk *Momente* (Hueber Verlag) (Original in Farbe)

Die von mir im Unterricht am häufigsten eingesetzten Tools sind *Padlet* und der *Bookcreator*<sup>11</sup>, bei dem die Teilnehmenden wieder gemeinsam etwas erarbeiten. Bei *Padlet* gebe ich den Teilnehmenden jeweils den Link und alle können zeitgleich aktiv werden. Hier Beispiele meiner *Padlets*:



Abb. 10: Auswahl der *Padlets* von Marion Grein (Original in Farbe)

Beim *Bookcreator* startet einer der Teilnehmenden ein „neues“ Buch und teilt dann den Einladungscode mit den Mitgliedern ihrer oder seiner Gruppe:

.....  
11 <https://bookcreator.com/>

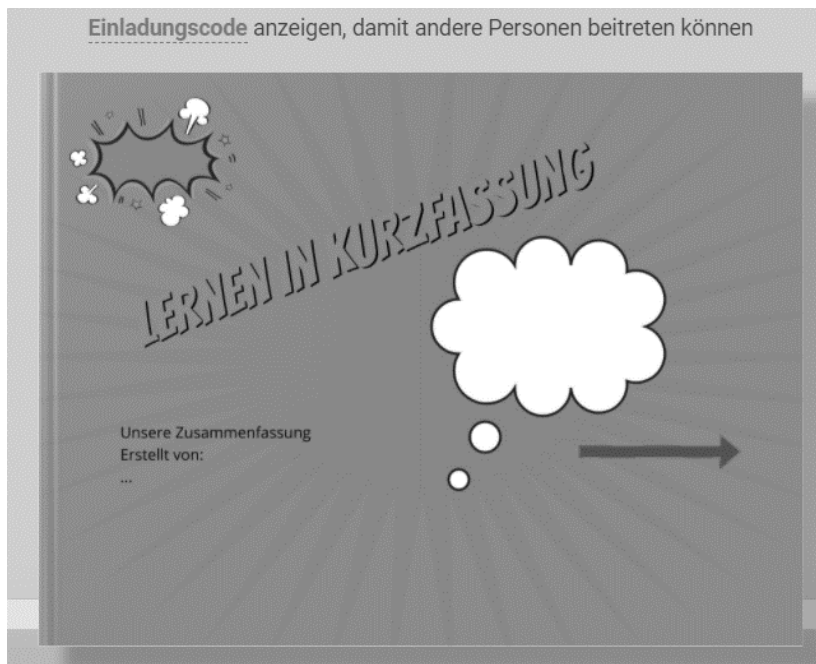


Abb. 11: Screenshot *Bookcreator* (Original in Farbe)

Mit dem *Bookcreator* erstellen die Lernenden ein eigenes Produkt, das sie dann im Plenum den anderen vorstellen. Diese Beispiele wurden in den Seminaren/Workshops usw. vorgestellt und mit den Lehrenden diskutiert. Der Einsatz dieser Variante des virtuellen Sprachunterrichts wurde in der Studie und in der folgenden Darstellung als Typus 5 ausgemacht. Doch welche Formen des Sprachunterrichts ergaben sich aus den Gesprächen mit unterschiedlichen Institutionen?

### 3 Die unterschiedlichen Umsetzungen des virtuellen Sprachunterrichts

Bei den einzelnen Fortbildungen wurden die Sprachkursleitenden nach ihrer Umsetzung des virtuellen Sprachunterrichts gefragt. Dabei ergaben sich grob

fünf unterschiedliche Typen, die jedoch sicherlich durch weitere Typen ergänzt werden müssen und sich durch zahlreiche weitere Faktoren unterscheiden. Manche Institutionen haben dabei auch mehrere Varianten eingesetzt. Gebündelt ergeben sich die folgenden fünf Typen des virtuellen Sprachunterrichts:

**Typ 1:** Der Unterricht wird lediglich über ein Webchat-Tool so wie herkömmlich gestaltet, ohne Integration der digitalen Versionen der Lehrwerke und ohne den Einsatz von ergänzenden Tools. Die Lernenden haben ihre Lehrwerke vor sich liegen und die Lehrkraft versucht, die Teilnehmenden zur Mitarbeit und zu Aktivitäten anzuregen. Die Lehrkräfte berichten, dass das vor allem mit Lernenden, die sie bereits kennen, recht gut funktioniert.

**Typ 2:** Der Unterricht wird auch hier über ein Webchat-Tool angeboten, allerdings wird die digitale Variante des Lehrwerks genutzt, so dass Hörtexte und Filme leicht gezeigt werden können. Weitere Tools kommen hier nicht zum Einsatz.

**Typ 3:** Manche Institutionen haben in der ersten Phase auf vorhandene Lernplattformen zurückgegriffen und auf den Einsatz von Lehrwerken verzichtet. Das *vhs-Portal* diente dabei ursprünglich als Ergänzung zur Präsenzlehre, wurde aber dann auch als Plattform zum selbstständigen Lernen angeboten. In meiner Evaluation des Lernportals (Grein 2020c) wurden die positiven Eigenschaften bereits hervorgehoben, allerdings kam ich dort zu dem Schluss, dass es sich perfekt zur Ergänzung der Lehre anbietet, aber nicht als Plattform zum selbstständigen Lernen (vgl. Grein 2020c: 14). Positiv bewertet wurden die Möglichkeiten sich über einen Chat (nur schriftlich) auszutauschen, mit Foren zu arbeiten und das Tutorensystem. Hinzugekommen ist eine Pinnwand. Inzwischen gibt es auch Volkshochschulen, die das *vhs-Portal* mit den Lehrwerken der drei Verlage *Hueber*, *Klett*, *Cornelsen* kombinieren; das ist sicherlich eine ideale Lösung. Während *telc* m.E. hervorragende Prüfungen erstellt<sup>12</sup>, sind deren Lehrwerke nicht kompetenz-, sondern eher prüfungsorientiert (vgl. Tornero Pérez 2019: 557), so dass ich den Einsatz nicht empfehlen

.....

12 Vgl. Jehle in diesem Band.

würde. Laut Aussagen vieler Befragter wird jedoch häufig auf das Selbstlernen und die Korrekturen der Aufgaben zurückgegriffen. Diese Form verstehe ich in der Studie als Typ 3.

**Typ 4:** Parallel zu den Angeboten des *vhs-Lernportals* hat das Goethe-Institut eine von Lehrwerken unabhängige Lernplattform entwickelt, die über sehr viele reaktive Übungstypen, Hörtexte, Filme usw. verfügt und zum selbstständigen Lernen gedacht ist. Das Prinzip entspricht dabei dem des *Inverted Classroom*. Die Lernenden lernen selbstständig und haben dann regelmäßige Treffen mit den Lehrenden. Die Gespräche mit einigen Lehrkräften waren sehr heterogen: Manche argumentierten PRO aufgrund der erhöhten Lernendenautonomie und der Flexibilität bezüglich der Uhrzeit, zu der die Teilnehmenden an den Online-Kursen teilnehmen können. Der Mehrheit fehlte jedoch der soziale Austausch, vor allem unter den Kursteilnehmenden selbst. Die Inhalte der Lernplattform sind sehr ansprechend, integrieren jedoch keine *EduApps*. Probleme werden auch mit dem Konferenztool *Adobe Connect* geäußert. Lediglich eine Lehrkraft ergänzte die Plattform *DUO*, die eine Zwischenform zwischen Typ 3 und 4 darstellt. *DUO* richtet sich an zukünftige Studierende und arbeitet mit einem Tutorensystem, in dem gerade bei Anfängern und Anfängerinnen sehr viel Wert auf *Voice-Chats* gelegt wird. Die Materialien finden sich alle auf der Lernplattform und das Selbstlernen steht im Vordergrund. Es werden jedoch unterschiedliche Varianten angeboten (siehe <https://www.deutsch-uni.com/de/>; 25.10.2020).

**Typ 5:** Hier wird das von mir in Kapitel 2 dargestellte Konzept verwendet, also digitales Lehrwerk und die Integration durch zahlreiche *EduApps*, die auch den sozialen Austausch fördern.

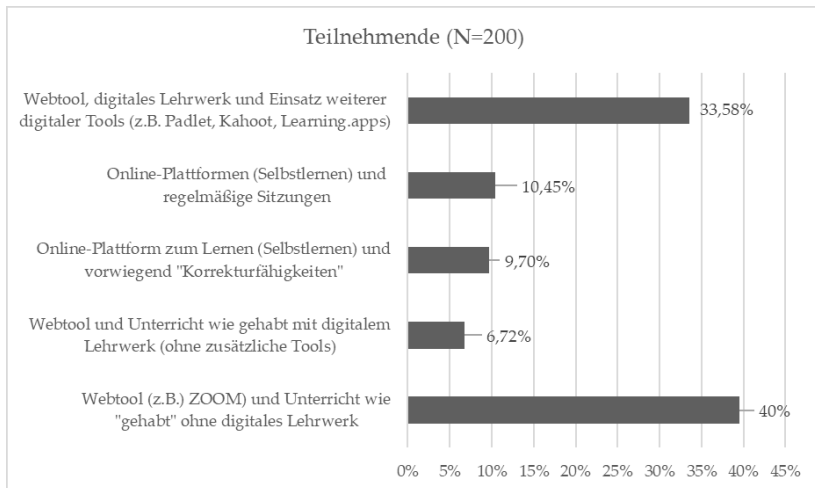
## 4 Die kurze Studie

Um zu quantifizieren, welche der fünf genannten Typen wie oft zum Einsatz kommt, wurde eine sehr kurze Online-Umfrage über *SurveyMonkey* (Oktober 2020) durchgeführt. Die Ergebnisse zeigen lediglich Tendenzen, und es bedarf



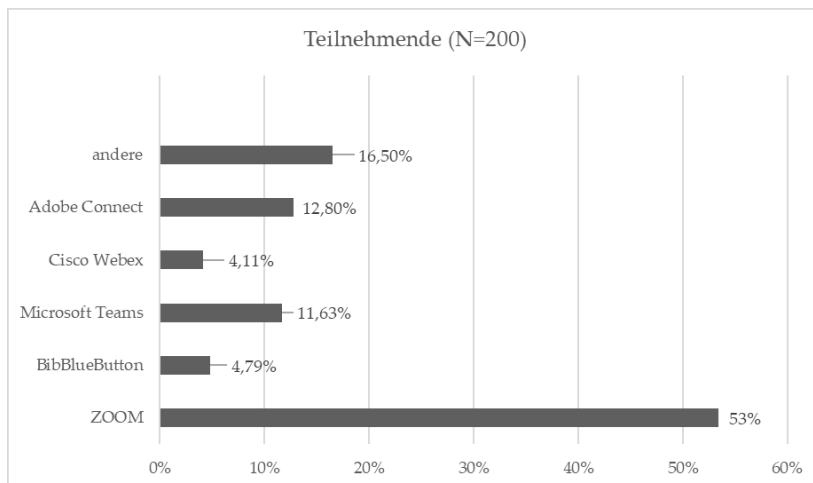
zahlreicher weiterer, vor allem auch qualitativer Studien. Zahlreiche Fragen werden in den zum Wintersemester 2020/21 angemeldeten Masterarbeiten aufgegriffen. Derzeit haben 200 Lehrende an der Umfrage teilgenommen, die Befragung bleibt jedoch für weitere Ergebnisse auch nach Veröffentlichung des Artikels offen, um die Daten zu validieren.

Zunächst wurde nach dem (vorwiegenden) Typus des virtuellen Unterrichts gefragt. Hier gab es Rückmeldungen, dass manche Lehrende an unterschiedlichen Institutionen mit unterschiedlichen Typen arbeiten. Diese eignen sich daher dann besonders für eine qualitative Befragung. Wie sah nun die Verteilung aus?



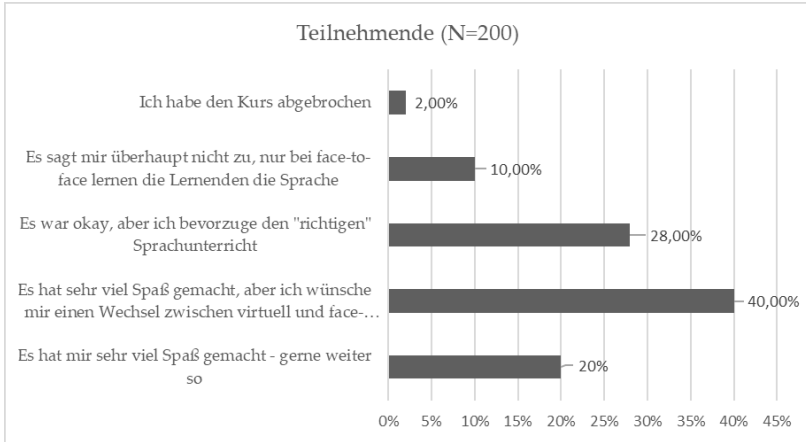
Immerhin etwas mehr als ein Drittel der Lehrenden haben folglich ihren virtuellen Unterricht so konzipiert, wie in meinen Weiterbildungen und Fortbildungen favorisiert (Typ 5). Ca. 20% haben die Selbstlernvarianten gewählt oder wählen müssen, da von der Institution vorgegeben (entspricht Typen 3 & 4).

Die zweite Frage bezog sich auf das eingesetzte Video-Chat Tool und zeigte, dass das Konferenztool Zoom zweifelsohne an erster Stelle steht.



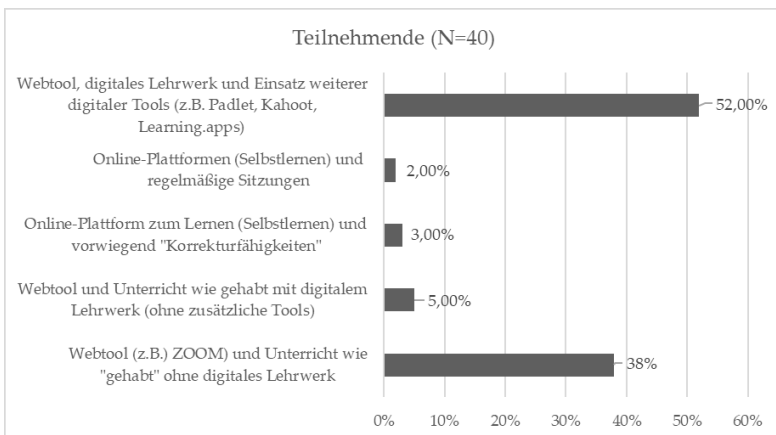
Bei den Kommentaren findet sich wiederum der Hinweis, dass manche Lehrenden mit mehreren Video-Chat Tools gearbeitet haben. Nicht aufgeführt wurden *GoToMeeting*, *EduDip*, *Jitsi*, da nur von wenigen genutzt. Ergänzt wurden *Vitero*, *Google Meet*, *Skype*, *Collaborate*, und *Moodle*, wobei *Moodle* ein LMS und kein Webchat-Tool ist. Es ist daher zu vermuten, dass hier *BigBlueButton* nicht aktiviert war.

Spannend nun die 3. Frage, bei der die Teilnehmenden eine kurze Gesamtevaluation abgeben, die es dann mit den eingesetzten Typen zu korrelieren gilt. Zunächst das Gesamtbild:



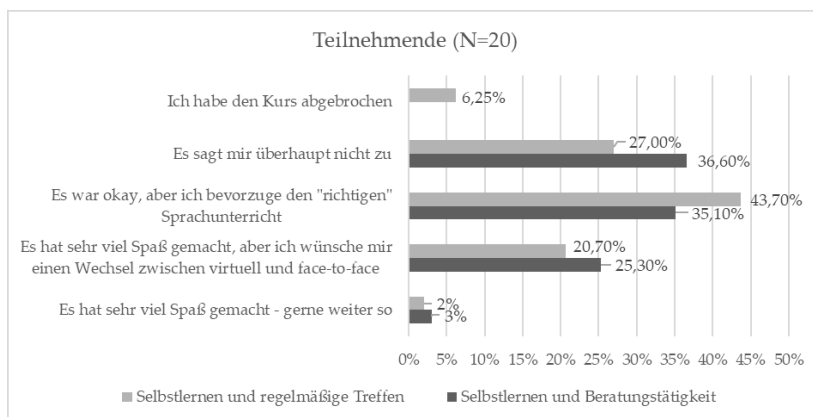
Immerhin 20% der Befragten können sich gut vorstellen, ihren Fremdsprachenunterricht auch in Zukunft virtuell anzubieten. Lediglich ca. 10,2% konnten sich mit dem Online-Unterricht gar nicht anfreunden. Sehr spannend und sicherlich zukunftsweisend sind die 40%, die sich in der Zukunft ein hybrides Konzept vorstellen können, also den Wechsel zwischen *face-to-face* und virtuellem Sprachunterricht.

Die Aussage „es hat sehr viel Spaß gemacht – gerne weiter so“ (N=40) verteilt sich wie folgt auf die fünf Typen:



Bei sowohl Typus 1 (38%) als auch Typus 5 (52%) können sich einige die virtuelle Lehre auch in der Zukunft vorstellen; bei den Selbstlern-Plattformen sind es deutlich weniger (5%), die sich damit anfreunden können. Hier zeigt sich vor allem der Bedarf an weiteren Studien, in denen erhoben wird, wie der Unterricht in Typus 1 umgesetzt wurde, denn auch hier sind viele Lehrende sehr zufrieden mit ihrem Online-Sprachkurs. Einer genaueren Analyse bedarf auch das Ergebnis, dass die reinen Selbstlernkurse als weniger erfolgreich eingestuft wurden. Was genau sind die Gründe?

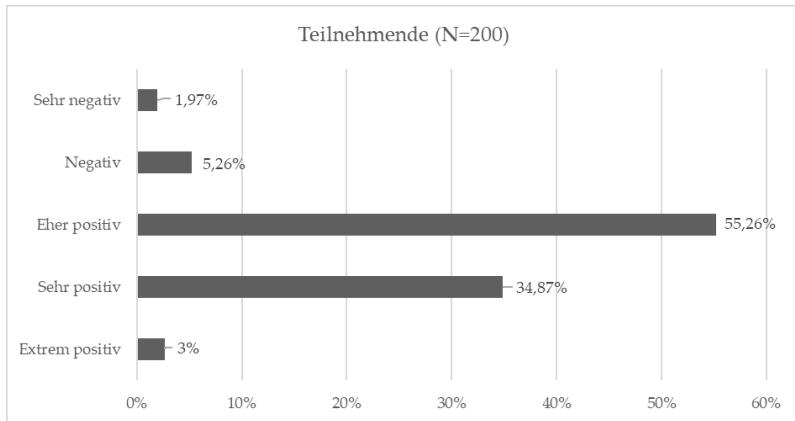
Korrelieren wir noch die beiden Selbstlernkurse (Typ 3 und Typ 4) mit der Frage 3:



Offensichtlich hat den Lehrenden das Konzept „Selbstlernen mit den regelmäßigen Treffen“ (Typ 4) am wenigsten Spaß gemacht. Es gilt also in einer weiteren Studie zu erheben, warum dieser Typus am wenigsten positiv bewertet wurde. Zwei Lehrende haben den Kurs sogar abgebrochen.

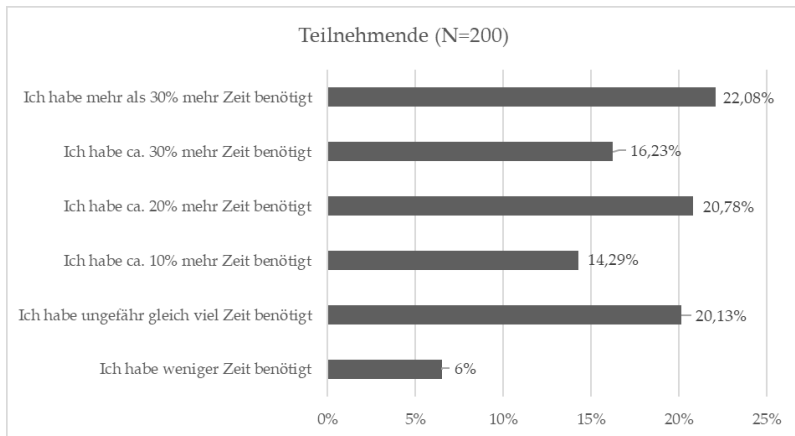
Gefragt wurden die Teilnehmenden dann auch, ob sie annehmen, dass den Lernenden der Unterricht gefallen hat. Betrachten wir zunächst wieder den Gesamtüberblick:

### Wie denken Sie würden Ihre Teilnehmenden den Kurs bewerten?



Die Lehrenden sind also im Durchschnitt zufrieden mit Ihrem Sprachkurs. Die wenigen negativen Einschätzungen sind wiederum den Selbstlernkursen zuzuordnen.

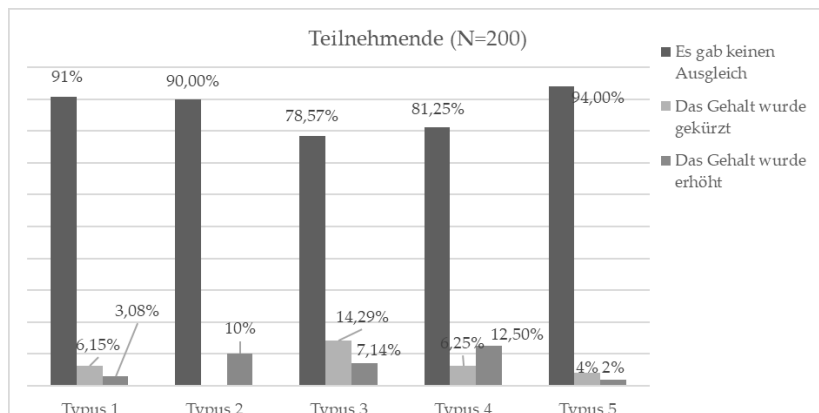
Es wurden zwei weitere Fragen – auf Wunsch auch der Befragten – hinzugefügt: Zum einen die Frage nach der Vorbereitungszeit, zum anderen die Frage nach dem finanziellen Ausgleich bei größerer Vorbereitungszeit.



Eine Korrelation zum Kurstyp ist hier nicht auszumachen.

Deutlich wird jedoch, dass die meisten Lehrenden einen erheblichen Mehraufwand bei der Vorbereitung der Kurse hatten, der jedoch sicherlich, wird der gleiche Kurs noch einmal angeboten, geringer ausfallen wird. Trotz des erhöhten Aufwands gab es in den meisten Fällen keinen finanziellen Ausgleich.

Hierbei ergibt sich folgende Verteilung:



## 5 Zusammenfassung und Kritik

Ziel der sehr kleinen Prästudie war es, einen Überblick über die derzeit eingesetzten virtuellen Sprachkurse zu erhalten. Der in Kapitel 2 dargestellte Ansatz, also der Typus 5, wurde dabei etwas besser bewertet als die anderen Typen, wobei auch der Typus 1 sehr viel erfolgreicher bewertet wurde als angenommen. Die reinen Online-Kurse (Selbstlernkurse) schneiden insgesamt am schlechtesten ab. Diese Prästudie hat zahlreiche Schwachpunkte, die in Folgestudien ausgeglichen werden müssen. Allem voran wurde nicht nach Kursform (DaF/DaZ), Niveaustufe und Jahren der Lehrerfahrung differenziert. Auch die einzelnen Typen müssen in Folgestudien sehr viel ausführlicher beschrieben und stärker differenziert werden. Grundsätzlich zeigt sich jedoch die Tendenz, dass der Sprachunterricht in Zukunft stärker hybrid gestaltet werden kann und den

Institutionen bewusst werden muss, dass qualitativ guter Online-Unterricht mehr Vorbereitung benötigt als der herkömmliche *face-to-face*-Unterricht, auch wenn die Vorbereitungszeit nach einer gewissen Eingewöhnungsphase wahrscheinlich etwas geringer werden wird.

## Bibliografie

- ALEXIOU-RAY, J. & BENTLEY, C.C. (2016): „Faculty Professional Development for Quality Online Teaching“, in: *Journal of Distance Learning Administration* 18/4: 1–16.
- APPEL, M. & SCHREINER, C. (2014): „Digitale Demenz? Mythen und wissenschaftliche Befunde zur Auswirkung von Internetnutzung“, in: *Psychologische Rundschau* 65/1, Göttingen: Hogrefe Verlag: 1–10.
- DE FLORIO-HANSEN, I. (2018): „Lehren und Lernen von Fremdsprachen im Kontext der Digitalisierung“, in: MARTINEZ, H. & MEISSNER, F.-J. (Hrsg.): *Fremdsprachenunterricht in Geschichte und Gegenwart: Festschrift für Marcus Reinfried* (Giessener Beiträge zur Fremdsprachendidaktik), Tübingen: Narr: 291–302.
- EICKELMANN, B., LORENZ, R. & ENDBERG, M. (2016): „Die eingeschätzte Relevanz der Phasen der Lehrerausbildung hinsichtlich der Vermittlung didaktischer und methodischer Kompetenzen von Lehrpersonen für den schulischen Einsatz digitaler Medien in Deutschland und im Bundesländervergleich“, in: BOS, W. et al. (Hrsg.): *Schule digital – der Länderindikator 2016. Kompetenzen von Lehrpersonen der Sekundarstufe I im Umgang mit digitalen Medien im Bundesländervergleich*, Münster: Waxmann: 149–182.
- FUNK, H. (2019): „Feindliche Übernahme oder erweiterte didaktisch-methodische Szenarien? Fremdsprachenunterricht in Zeiten des digitalen Wandels“, in: BURWITZ-MELZER, E. RIEMER, C. & SCHMELTER, L. (Hrsg.): *Das Lehren und Lernen von Fremd- und Zweitsprachen im digitalen Wandel*, (Giessener Beiträge zur Fremdsprachendidaktik), Tübingen: Narr: 68–80.

- GREIN, M. & STRASSER, T. (2019): „Lernen mit digitalen Medien aus neurobiologischer und fremdsprachendidaktischer Sicht“, in: *Empfehlungen Goethe-Institut Zagreb*: 7–15.
- GREIN, M. (2020): „Sprachunterricht auf einmal digital: Online-Lernen vor, während und nach Corona“, in: *Magazin Sprache*, Goethe Institut, o.S., abrufbar unter: <https://www.goethe.de/de/spr/mag/21927962.html>, 09.09.2020.
- GREIN, M. (2020b): „Digitalen Sprachunterricht interaktiv gestalten: nützliche Tools und Tipps zur Anwendung“, in: *Magazin Sprache*, Goethe Institut, o.S., abrufbar unter: <https://www.goethe.de/de/spr/mag/21927962.html>, 09.09.2020.
- GREIN, M. (2020c): „Über das Lernen mit digitalen Medien (Interview)“, in: *vhs dis.kurs Ausgabe 1/2020*, Bonn: Deutscher Volkshochschulverband e.V.
- LORENZ, R. & ENDBERG, M. (2019): „Welche professionellen Handlungskompetenzen benötigen Lehrpersonen im Kontext der Digitalisierung in der Schule? Theoretische Diskussion unter Berücksichtigung der Perspektive Lehramtsstudierender“, in: *Medien Pädagogik. Zeitschrift für Theorie und Praxisbildung der Medienbildung* (Occasional Papers): 61–81, abrufbar unter: <https://doi.org/10.21240/mpaed/00/2019.10.16.X>.
- PERSIKE, M. & FRIEDRICH, J.-D. (2016): „Lernen mit digitalen Medien aus Studierendenperspektive“, in: *Arbeitspapier Nr. 17*, Berlin: Hochschulforum Digitalisierung.
- ROCHE, J. (Hrsg.) (2019): *Medienwissenschaft und Mediendidaktik*, Tübingen: Narr.
- RÖSLER, D. (2019): „Auf dem Weg zum Babelfisch? Fremdsprachenlernen im Zeitalter von Big Data“, in: *Informationen Deutsch als Fremdsprache* (published online ahead of print 2019), abrufbar unter: <https://doi.org/10.1515/infodaf-2020-0060>.
- SCHIEFNER-ROHS, M. & Autorengruppe (2018): „Lehrerbildung für die digitale Zukunft“, in: *Synergie. Fachmagazin für Digitalisierung in der Lehre* 6: 48–55.
- SCHMIDT, T. & STRASSER, T. (2016): „Digital Classroom“, in: *Der fremdsprachliche Unterricht Englisch* 50: 2–7.



TORNERO PÉREZ, M. (2019): „Rezension zu Einfach gut! Deutsch für die Integration. Kurs- und Arbeitsbuch A1–B1 sowie Lehrerhandbuch A1–B1“, in *ZIF 24/2*, abrufbar unter: <https://tjournals.ulb.tu-darmstadt.de/index.php/zif/article/view/999/996>, 25.10.2020.

VAN ACKEREN, I. et al. (2019): „Digitalisierung in der Lehrerbildung. Herausforderungen, Entwicklungsfelder und Förderung von Gesamtkonzepten“, in: *Die Deutsche Schule 111/1*: 103–119, abrufbar unter: <https://doi.org/10.31244/dds.2019.01.10>; 25.10.2020.

## Lehrwerk

EVANS, S., PUDE, A. & SPECHT, F. (2020): *Momente*, München: Hueber Verlag.

## **Learn to flirt by *Duolingo* „Du siehst aus wie meine nächste Freundin“ – Die Evaluation von Sprachlern-Apps für Deutsch als Fremdsprache in Bezug auf die Vermittlung interkultureller Kompetenz**

### **Einleitung**

Nutzen Lernende schon seit geraumer Zeit digitale Anwendungen zum Erlernen einer Sprache, da sich mit dem Lernen durch Apps auch „empty‘ time slots during the day“ (Falk & Götz 2016: 238) füllen lassen, wurde das Thema „Sprachlern-App“ im Frühjahr des Jahres 2020 wohl aktueller denn je. Durch Ausgangsbeschränkungen und Lockdowns vieler Länder verzeichneten Sprachlern-Apps wie etwa *Duolingo* einen deutlichen Zuwachs an Nutzer/innen, so zum Beispiel allein im März 2020 von 108% (Ward 2020). Eine der meistgewählten Sprachen bei Nutzenden der englischen Version der Applikation war in diesem Zeitraum „German for English Speakers“, welche von 8,18 Millionen Menschen weltweit installiert wurde (ebd.) Zu dieser und anderen Online-Anwendungen wurden bereits verschiedene Umfragen, Analysen und Evaluationen durchgeführt (vgl. dazu Falk & Götz 2016, Usta 2017), welche die App, deren Funktionen und die Motivation der Lernenden zur Verwendung dieser Anwendung untersuchten. Laut Godwin-Jones fokussieren viele Studien zum Thema jedoch „only institutional use of mobile devices (i.e., in formal, instructed language learning). Few examine use out-side the classroom“ (2017: 5). In Zeiten, in denen das Erlernen einer Fremdsprache jedoch im Klassenraum nicht unbedingt möglich ist, bleiben Lernenden Sprachlern-Apps zum autonomen Sprachenlernen. Hierbei stellt sich die Frage, ob Lernende mit Satzbausteinen wie den im Titel genannten „Du siehst aus wie meine nächste Freundin“ aus der App *Duolingo* (Learn to flirt in German) jedoch wirklich auf den realen, kulturbezogenen Sprachgebrauch vorbereitet werden. Denn

neben den zu erlernenden Fertigkeiten Lesen, Hören, Schreiben und Sprechen ist „interkulturelle Kompetenz, verstanden als grundlegende Handlungskompetenz, für jeden Menschen essenziell“ (Grein 2019: 125). So müssen Lernende interkulturell kompetent sein, um mit Muttersprachler/innen des Deutschen erfolgreich zu kommunizieren oder sich auf einen Aufenthalt im deutschsprachigen Raum vorbereiten zu können, da nur so eine erfolgreiche Interaktion zwischen Menschen verschiedener Kulturen gelingen kann (vgl. Fuchs 2012: 17). „[N]icht alle Teilkonstrukte [der interkulturellen Kompetenz, A-K.F.] sind vermittelbar. Voraussetzung ist jedoch die Sensibilisierung für kulturspezifische Unterschiede“ (Grein 2019: 128); diese mit in den Fremdsprachenunterricht sowie die hierfür verwendeten Lehrwerke mit einzubeziehen ist der Anspruch des modernen Fremdsprachenunterrichts. Auch Sprachlern-Apps erheben den Anspruch, ihren Nutzer/innen eine Sprache so zu vermitteln, dass „echte Kommunikation“ (Babbel 2020: o.S.) möglich ist. Können Sprachlern-Apps hier jedoch als Ersatz echten DaF-Unterrichts dienen, die nötigen Kompetenzen den autonom Lernenden ausreichend vermitteln und sie somit nicht nur in den vier Fertigkeiten, sondern auch für interkulturelle Themen sensibilisieren?

Dieser Artikel beschäftigt sich mit der Frage, ob und wie interkulturelle Kompetenz mit digitalen Anwendungen aufgegriffen wird und inwieweit Lernende dadurch für kulturelle Unterschiede sensibilisiert werden können. Hierfür werde ich drei der meist heruntergeladenen Apps – mit jeweils über 10 Millionen Downloads im *Google PlayStore* (vgl. [play.google.com](https://play.google.com)) – *Babbel*, *Duolingo* und *Busuu*, auf interkulturelle Elemente und die Sensibilisierung und Förderung der interkulturellen Kompetenz ihrer Nutzer/innen untersuchen. Denn nur mit dieser lassen sich Sprachlern-Apps zumindest als kurzweiliger Ersatz für den interkulturellen Fremdsprachenunterricht einsetzen. In Kapitel 1 wird somit zunächst kurz auf allgemeine Sprachlern-Apps eingegangen, gefolgt vom zweiten Kapitel, das die interkulturelle Kompetenz und deren Rolle im Fremdsprachenunterricht fokussiert und eine Einschätzung zu erwartender interkultureller Komponenten in Sprachlern-Apps gibt. Im empirischen Teil (Kapitel 3) werden die beschriebenen Apps auf diese Komponenten hin untersucht und die Ergebnisse anschließend zusammengefasst, bevor ein Ausblick für den DaF-Unterricht gegeben wird.

## 1 Sprachlern-Apps

Durch eine immer digitaler werdende Welt hat sich auch das Sprachenlernen längst im digitalen Bereich verfestigt und macht das so genannte *Mobile Learning* zum Alltag vieler Lernender: Während man auf einen Bus wartet, eine kurze Pause hat oder noch fünf Minuten Zeit bleiben, werden diese nicht nur dafür genutzt, sich mit Freunden auszutauschen oder Nachrichten zu lesen, sondern auch zum Sprachen Lernen (vgl. Falk & Götz 2016: 238). Der Begriff des *M-Learnings* beschreibt eben diese Flexibilität, die das Lernen mit einem mobilen Endgerät überall und zu jeder Zeit möglich macht und „betont [...] die räumliche Trennung von Lernen und dem Klassenzimmer“ (Usta 2017: 11). Ein wichtiger Punkt des *M-Learnings* ist die Dauer der Lerneinheit, welche im Gegensatz zum festgelegten Unterricht kürzer und variabler ist (vgl. ebd., ebenso Lai & Zheng 2017: 307). So bietet *M-Learning* „not only the mobility of technology but also more importantly the mobility of time, space, and learning experience“ (Lai & Zheng 2017: 300) und unterscheidet sich somit vom erweiterten Begriff des E-Learnings, welcher das Lernen mit digitalen Medien allgemein verknüpft. „Wenn man sich vor den Computer setzt, um mit einer Lernsoftware zu lernen, plant man ebenfalls eine bestimmte, evtl. längere, Zeit zum Lernen ein“ (Usta 2017: 11), beim *M-Learning* ist der/die Lernende flexibler, wobei das Konzept sich nicht auf Smartphones beschränkt, sondern beispielsweise auch Tablets hierfür genutzt werden können. Nutzen Lernende *M-Learning*, also z.B. Sprachlern-Apps, um Vokabeln zu lernen (vgl. Lai & Zheng 2017: 305) oder um die verschiedenen Fertigkeiten zu üben (vgl. Usta 2017: 58), besteht bei einem Vergleich verschiedener Apps eine Schwierigkeit: „Apps vary tremendously in purpose, scope, and design, and they need to be judged individually“ (Godwin-Jones 2017: 6). Trotz dieser Varietäten versprechen die meisten das Sprachenlernen dennoch „immer unter Berücksichtigung realistischer Szenarien“ (*Babbel*) und, dass Sprache als „Kommunikationsinstrument“ vermittelt wird, Lernende also „sehr schnell dazu bringt, [...] eine neue Sprache wirklich zu sprechen. Denn das ist die größte Motivation für jeden Lernenden“ (*Babbel*). Um dies zu erreichen, also eine Kommunikation mit Muttersprachler/innen erfolgreich durchzuführen, bedarf es nicht nur Vokabular und Grammatik, sondern auch der interkul-

turellen Kompetenz der Lernenden, auf die in den folgenden Abschnitten eingegangen wird.

## 2 Interkulturelle Kompetenz und ihre Vermittlung

Gilt die interkulturelle Kompetenz heutzutage als „die Fähigkeit individuelle, soziale, fachliche und strategische Teilkompetenzen in ihrer bestmöglichen Verknüpfung auf interkulturelle Handlungskontexte beziehen zu können“ (Bolten 2006, zit. in Grein 2019: 129), galt ihre Definition und Position im Fremdsprachenunterricht nicht von je her als gegeben, sondern bildete sich erst Anfang der 1980er Jahre als Unterrichtsziel heraus (vgl. Fuchs 2016: 18).

Doch auch schon bevor die interkulturelle Kompetenz als übergreifendes Ziel des Fremdsprachenunterrichts angesehen wurde, wurde Kulturwissen im Teilgebiet der Landeskunde vermittelt. Im kommunikativen Ansatz wollte man in der Landeskunde weitestgehend auf Fakten verzichten und sich auf alltägliche, eben kommunikative Situationen konzentrieren (vgl. Rösler 2012: 203). Galt die Landeskunde somit zunächst als Mittel zum Zweck, also zur Themenfindung von Sprechansätzen, entwickelte sie sich in der Phase der problemorientierten Landeskunde zu einem Rahmen für interkulturelle Vergleiche. Daraus resultierten seit den 1980ern die ersten Versuche, eine interkulturelle Landeskunde zu erreichen. Hierbei liegt der Hauptfokus laut Rösler (2012: 205) nicht auf „der Ebene des Wissens“ sondern auf „der Ebene der Einstellung“ zu Zielkultur und -sprache. Das Unterrichtsziel der heutigen interkulturellen Kompetenz erweitert das Konzept und fokussiert nicht nur die Zielkultur selbst, sondern bezieht die Reflexion der eigenen Kultur mit ein (vgl. Usener 2016: 7). Dennoch bleibt die Landeskunde selbst bedeutsam für die interkulturelle Kompetenz, da sie als „erster Schritt des interkulturellen Lernens“ gesehen werden kann (ebd.: 8). Cataldo (2017: 9) beschreibt die Entwicklung von Landeskunde zur interkulturellen Kompetenz als

eine Bewegung, die sich immer konsequenter von der bloßen Vermittlung von Faktenwissen über die fremde Kultur entfernt und zu einer erweiterten Kommunikations- und Verstehensbereitschaft im inter-

kulturellen Kontext hinarbeitet. Hier werden Sprache und Kultur als untrennbar verstanden. Die Konsequenz daraus ist die IK als übergreifendes Kompetenzziel des [Fremdsprachenunterrichts, A.-K. F].

Da das Feld der interkulturellen Kompetenz sowie deren Vermittlung ein sehr komplexes ist, wurde seit den 1980er Jahren versucht, das Konstrukt aufzuteilen und zu veranschaulichen. Grein (2018: 126) fasst zusammen, dass hier zunächst Teilkompetenzen wie „empathy“, „cultural awareness“ oder „respect for cultural differences“ herausgearbeitet wurden, während sich die Tendenz in der Forschung unter Berücksichtigung dieser Konzepte zu Strukturmodellen verlagerte. Byram (1997) entwickelte beispielsweise eine Teilung in fünf Kompetenzen, sogenannte „savoir“, die aus Fachwissen, Einstellungen des Lernenden, dessen Fähigkeit, eine andere Kultur zu verstehen, neues Wissen zu erwerben und eine fremde Kultur kritisch zu reflektieren, bestehen (vgl. Cataldo 2017: 12, ebenso Usener 2016: 24). „Allerdings gehen einzelne Teilkompetenzen fließend ineinander über und sind nicht immer eindeutig voneinander abzugrenzen“ (Usener 2016: 24); eine Problematik, die sich auch auf andere Modelle übertragen lässt, welche die interkulturelle Kompetenz in kognitive (Wissen um kulturspezifische Unterschiede), affektive (Einfühlungsvermögen) und verhaltensbezogene (Respekt vor kulturspezifischen Unterschieden) einteilt (vgl. Grein 2019: 127). Auch wenn die Komponenten laut Grein (ebd.) von Müller & Gelbrich (2014) um Effektivität (Anpassungsfähigkeit) und Angemessenheit (Befolgung von Regeln, Anerkennung von Gefühlen) ergänzt wurden, bleibt bei diesem strukturellen Modell zum einen das Problem der Abgrenzung, aber auch, dass die Teilgebiete nicht „getrennt voneinander vermittelbar“ sind. Aus dieser Problematik folgt, dass heutzutage unter Berücksichtigung der drei Komponenten (affektiv, kognitiv, verhaltensbezogen) mehr prozessorientierte Modelle verwendet werden (vgl. ebd.).

Als Voraussetzung zum Erwerb interkultureller Kompetenz wird hierbei das Bewusstsein des eigenen Ethnozentrismus gesehen, dass also Menschen die eigene Kultur als Normalität bewerten:

Die Entwicklung eines neuen, eines reflektierten Ethnozentrismus ist deshalb das Ziel eines jeden interkulturellen Lernens und Trainings.

Irritationen und Fremdheitserfahrungen werden idealerweise kontrolliert wahrgenommen und reflektiert, Abwertungen werden vermieden oder kontrolliert wahrgenommen und die eigenen Reaktionen werden in Selbstreflexion hinterfragt (Witzenleiter & Luppold 2020: 46).

Um interkulturell kompetent agieren zu können, wird außer diesem Bewusstsein jedoch auch Wissen darüber vermittelt, in welchen Bereichen sich Kulturen – ohne negative oder positive Bewertung dieser – unterscheiden können und eine damit verbundene „Sensibilisierung für kulturspezifische Unterschiede“ erreicht (Grein 2019: 128). Hierbei darf auch die Individualität eines Menschen nicht außer Acht gelassen werden (vgl. ebd.). So darf es zu keinem stereotypischen Kulturverständnis kommen, bei dem davon ausgegangen wird, dass alle Mitglieder einer Kultur ‚gleich‘ sind. Diese Idee findet sich als zentraler Punkt des Multikollektivismus wieder, der besagt, dass Menschen eines Kulturraums zu verschiedenen Kollektiven gehören können und diese verschiedenen Zugehörigkeiten eine Person ausmachen (vgl. ebd.). Dies stellen auch Witzenleiter & Luppold fest: „Eine Begegnung eines im Dorf lebenden Studiengangleiters mit einem Mitglied des Schützenvereins kann eine interkulturelle Erfahrung sein. Sobald Menschen sich gegenseitig als ‚fremd‘ erleben, findet ‚Interkulturelle Kommunikation‘ statt“ (2020: 4).

Kann eine Sprachlern-App jedoch gewährleisten, dass die erlernbaren Teilbereiche der interkulturellen Kompetenz abgedeckt und vermittelt werden? Ein Merkmal des *M-Learnings*, welches zuvor erläutert wurde, ist das eigenständige und ortsunabhängige Lernen: So sind die Nutzer der App, die ohne Lehrkraft oder andere Mitlernendem arbeiten, mit denen interkulturelle Kompetenz vermittelt und thematisiert werden kann, auf sich alleine gestellt. Hierbei stellt sich also die Frage, durch welche Mittel eine Applikation Lernende für interkulturelle Unterschiede sensibilisieren kann und ob diese auch ohne den entsprechenden Austausch mit einer Lehrkraft oder anderen Lernenden ausreichend reflektiert wird. Um dieser Frage weiter nachzugehen, wird im folgenden Abschnitt die bisherige Forschung zur interkulturellen Kompetenzvermittlung in Apps skizziert.

## 2.1 Interkulturelle Kompetenz in Sprachlern-Apps

Mit der Popularität von Apps zum Fremdspracherwerb wächst auch die Forschungsarbeit auf diesem Gebiet. Rosell-Aguilar (2018: 5) weist jedoch hierbei daraufhin, dass sich die meisten Studien auf die Verwendung von Apps im oder neben dem Spracherwerb im Unterricht beziehen. Auf der Suche nach konkreten Studien oder Leitfragen zur Evaluation von Sprachlern-Apps außerhalb von Unterrichtssituationen findet man hauptsächlich in der englischsprachigen Literatur einige Ansätze (vgl. Rosell-Aguilar 2017, Chen 2016). Hierbei gehen Wissenschaftler/innen auf verschiedene Aspekte der Apps ein. So werden zum Beispiel das Feedback und die Motivation für Lernende durch diese Anwendungen untersucht (so in Chen 2016) oder aber die Nutzung digitaler Anwendung für die Verbesserung bzw. Korrektur verschiedener Fertigkeiten (vgl. Lai & Zheng 2017).

Rosell-Aguilar (2017) bietet hierfür einen detaillierten Kriterienkatalog zur umfassenderen Analyse und Evaluation von Apps. Hierbei geht er auf folgende Faktoren ein: Technology, Pedagogy, Language Learning und User Experience (vgl. ebd.: 14). Für den vorliegenden Beitrag interessant ist vor allem die Kategorie Language Learning, die in Abb. 1 dargestellt wird:

Language learning	
<ul style="list-style-type: none"> <li>▪ <b>Reading:</b> does the app provide texts in the target language?</li> <li>▪ <b>Listening:</b> does the app provide audio in the target language?</li> <li>▪ <b>Writing:</b> does the app offer opportunities to write in the target language?</li> <li>▪ <b>Speaking:</b> does the app offer opportunities to speak in the target language?</li> <li>▪ <b>Vocabulary:</b> does the app offer specific activities for vocabulary acquisition?</li> <li>▪ <b>Grammar:</b> does the app offer specific activities for grammar practise?</li> </ul>	<ul style="list-style-type: none"> <li>▪ <b>Pronunciation and intonation:</b> does the app offer specific activities for pronunciation and intonation?</li> <li>▪ <b>Cultural information:</b> does the app include information about customs and traditions in the areas where the language is spoken?</li> <li>▪ <b>Use of visual content:</b> are images and videos stereotypical or stock images? Do they represent the diversity of the areas where the language is spoken?</li> <li>▪ <b>Language varieties:</b> does the app include different regional or national varieties of the language?</li> </ul>

Abb. 1: Language Learning (Rosell-Aguilar 2017: 14)



Mit den Punkten „Cultural information“, „Use of visual content“ und „Language varieties“ geht der Autor auf Faktoren ein, die zum Erlangen der interkulturellen Kompetenz beitragen. Hierbei bieten die verwendeten Aspekte jedoch eine große Spannweite, da der Autor sich auf eine generelle Evaluation der App bezieht. Möchte man jedoch der Frage nachgehen, inwieweit interkulturelle Kompetenz vermittelt werden kann, so bietet es sich an, die Aspekte Rosell-Aguilars (2017) weiter zu vertiefen. Besonders im ersten Punkt „Cultural information“ wird sichtbar, dass der Autor auf die Informationsvermittlung eingeht, welche, wie in den vorherigen Abschnitten besprochen, vor allem der landeskundlichen Komponente der interkulturellen Kompetenz zugeschrieben werden kann. Der nächste Punkt „Use of visual content“ ist für die Untersuchung dahingehend interessant, da durch die Verwendung von Bildern und Videos eine Sensibilisierung der Lernenden stattfinden kann. Der Einsatz von Fotos und Bildern kann im Rahmen des Erlangens interkultureller Kompetenz verschiedene Funktionen haben. So eignen sich Bilder „aufgrund ihrer komprimierten Darstellung von Sichtweisen über die fremde und die eigene Kultur zur Sensibilisierung für wechselseitige Stereotype“ (Usener 2016: 48). Sie erlauben aber auch einen Einblick in das Alltagsleben der Zielkultur (vgl. ebd.: 51) und werden so „als Träger soziokulturellen Orientierungswissens“ wahrgenommen (vgl. ebd.: 96). Darüber hinaus können Bilder auch Identifikationsmöglichkeiten anbieten. Solche helfen den Lernenden „zu einer Identifikation mit fremdsprachlichen Unterrichtsinhalten, also u.a. zielkulturellen Werten und Verhaltensweisen“ (Hecke 2010: 100) zu gelangen. Durch diese Identifikation wird Empathie hervorgerufen, die die Lernenden dem interkulturellen Verständnis näher führt (vgl. ebd. 2010: 101). Dies geschieht, wenn es den Lernenden gelingt, sich in andere Rollen hineinzusetzen, sich der eigenen kulturellen Identität bewusst zu werden – somit haben Bilder ein hohes Emotionspotenzial (vgl. Biechele 2006: 20). Laut Usener (2016: 93) bietet es sich hierbei auch an, „den Lernern Figuren zu präsentieren, die auf authentischen Fotos abgebildet sind und demzufolge auch realiter existieren“. Sollen Bilder nicht explizit dafür genutzt werden, Stereotype zu thematisieren, ist es nach Hecke (2010: 93) außerdem ratsam, verschiedene Bilder zu benutzen, um die Entstehung von Stereotypen zu vermeiden. Biechele (2006: 32) regt im Zusammenhang von interkulturellen Kompetenzen

und Bildfunktionen zur Nutzung von authentischen Bildern an. Diese bieten eine „doppelte Offenheit“, eine, die auf der Polyvalenz von Bildern beruht, und eine, die der fremdkulturell geprägte Betrachter in seine Interpretation einbringt“ (ebd.: 36).

Auch mit der Frage „Does the app include different regional or national varieties of the language?“ geht Rosell-Aguilar (2017: 14) auf einen wichtigen Punkt der interkulturellen Kompetenz ein. Bei der Auswertung werden nicht nur Dialekte, sondern zum Beispiel auch Jugendsprache beachtet, welche Indikatoren für Multikollektivität sein können. Somit werden Lernende nicht nur mit den verschiedenen Sprachvarianten vertraut gemacht, sondern auch für Unterschiede sensibilisiert.

Werden von Rosell-Aguilar (2017) somit schon relevante Ansätze zur Evaluation berücksichtigt, bleibt jedoch die Frage offen, ob eine Sprachlern-App auch in anderen Bereichen zur Vermittlung interkultureller Kompetenz beitragen kann. Wie vom Autor festgestellt, können sich Apps in ihrer Gestalt sehr voneinander unterscheiden, sodass fraglich ist, welche Aspekte für die Sensibilisierung interkultureller Unterschiede erwartet werden können. Es bietet sich bei der Erstellung eines Fragenkatalogs also an, auch ähnliche bzw. bereits bekannte „Hilfsmittel“ des Sprachenlernens miteinzubeziehen und zu untersuchen, weshalb nachfolgend auf Elemente der Lehrwerksanalyse eingegangen wird.

## 2.2 Interkulturelle Kompetenz in Lehrwerken

Stellen Apps ein neueres Medium dar, welches zum Erlernen einer Fremdsprache genutzt wird, hat das Lernen mit analogem Lernmaterial hingegen eine lange Tradition (vgl. Ciepielewska-Kaczmarek et al. 2020: 19). Im Gegensatz zur App gibt es nicht nur in der Fremdsprachendidaktik allgemein, sondern auch speziell in Deutsch als Fremdsprache ein gut ausgebautes Forschungsgebiet zur Analyse von Lehrwerken. Auch die Landeskunde und die interkulturelle Kompetenz finden in diesem Feld Beachtung. Wie zuvor gezeigt, kann die Vermittlung von landeskundlichen Fakten als ein erster Schritt zum Erlangen interkultureller Kompetenz gesehen werden und stellt dabei einen Teilbereich der kognitiven Komponente dieser dar (vgl. Fuchs 2012: 50). Auch wenn

Sprachlern-Apps nur bedingt mit Lehrwerken aus dem Fremdsprachenunterricht vergleichbar sind, können die Kriterienkataloge, mit denen Lehrwerke analysiert werden, ebenfalls erste Anhaltspunkte für eine genauere Analyse zum *M-Learning* bieten.

Es finden sich verschiedene Ansätze, wie bspw. bei Ciepielewska-Kaczmarek et al. (2020), die verschiedene Lehrwerke anhand von Leitfragen zu landeskundlichen Inhalten analysieren. Hierbei untersuchen die Autorinnen zunächst, ob landeskundliche Themen explizit oder implizit behandelt werden bzw. ob diese als landeskundliche Themen ausgezeichnet sind (vgl. Ciepielewska-Kaczmarek et al. 2020: 77). Des Weiteren wird ausgewertet, ob Landeskunde additiv, isoliert bzw. integrativ dargeboten wird. Ist sie integrativ, wird sie mit sprachlichen Inhalten „verknüpft angeboten [...], wobei in den Aufgaben sprachliche und kulturelle Aspekte ‚gleichberechtigt‘ vertreten sind“ (ebd.). Als isoliert wird eine landeskundliche Einheit bezeichnet, wenn sie weder mit sprachlichen Aufgaben verbunden ist noch inhaltlich zur Einheit passt (vgl. ebd.). Als letzte Option sehen die Autorinnen die additive Landeskunde-Vermittlung, in der „ein vorgestelltes landeskundliches Phänomen zum Thema der Einheit passt, aber nicht mit sprachlichen Inhalten verbunden ist“ (ebd.). Ferner wird untersucht, inwieweit Bilder und Videos verwendet werden, ob ein Bezug zur Lebenswelt der Lernenden hergestellt wird und „ob die Aufgaben die Möglichkeit bieten, die ausgangs- und zielsprachige Umgebung zu entdecken und das Herstellen von Verbindungen zwischen beiden ermöglicht“, „eine kulturelle Haltung fördert“ und „den Lernenden ermöglichen, am zielsprachigen Diskurs teilzunehmen“ (ebd.: 78). Auch die Beachtung des DACH-Prinzips<sup>1</sup>, welches besagt, dass nicht nur Deutschland, sondern der weitere deutschsprachige Raum – hier Deutschland, die Schweiz und Österreich – und dessen kulturelle und sprachliche Vielfalt behandelt wird, wird untersucht (vgl. ebd.).

.....

1 Wird das DACHL-Konzept (also mit Berücksichtigung Liechtensteins) bei den Autoren angesprochen, liegt der Augenmerk der Untersuchung bei Ciepielewska-Kaczmarek u.a. 2020 (77) auf dem DACH-Konzept.

Liegt der Schwerpunkt dort auf landeskundlichen Elementen, geht User (2016: 89) auf die interkulturelle Kompetenz selbst ein und stellt einen ausformulierten detaillierten Kriterienkatalog zusammen, der 20 Fragen zur interkulturellen Kompetenz in Lehrwerken enthält:

#### **A Zur Stellung interkulturellen Lernens:**

1. Verweist das Inhaltsverzeichnis des Schülerbuchs bzw. der Lektionsauftrags auf interkulturelles Lernen als zentral zu erlernende Fertigkeit?
2. Weisen sowohl obligatorische als auch fakultative Lektionen und Lektionsteile interkulturelles Potenzial hinsichtlich ihrer Themen und Verfahren auf?
3. Werden Formen der Überprüfung interkultureller Kompetenz bzw. einzelner Teilkomponenten vorgeschlagen (z.B. Selbstevaluation)?

#### **B Zur Förderung der affektiv-attitudinalen Dimension (*savoir être*):**

4. Werden Lernerorientierung und Lebensweltbezug durch Themen-, Material- sowie Aufgaben- und Methodenauswahl deutlich?
5. Haben die Lernenden die Möglichkeit, sich mit den Lehrwerkfiguren zu identifizieren?
6. Werden Begegnungen mit der Fremdkultur bzw. Begegnungen einander fremder Kulturen thematisiert?
7. Werden critical incidents thematisiert und analysiert?
8. Werden stereotype Vorstellungen und Bilder gegenüber den Zielkultur(en) bzw. der Eigenkultur thematisiert und reflektiert?
9. Wird zur Hypothesenbildung zu Bildern und Texten sowie darin enthaltenen Handlungsabläufen und Verhaltensmustern angeregt?
10. Werden die Lernenden dazu angeregt, einen Perspektivwechsel zu vollziehen?

**C Zur Förderung der kognitiv-analytischen Dimension (*savoir*):**

11. Werden authentische Materialien (z.B. literarische Texte, Werbeanzeigen, Lieder, Fotos) als Wissenszugang zur fremdkulturellen Welt genutzt?
12. Wird soziokulturelles Orientierungswissen über Eigen- und Fremdkultur vermittelt bzw. thematisiert?
13. Wird allgemeines Wissen über Kultur und Kommunikation vermittelt?
14. Werden die kulturspezifische Prägung einzelner Begriffe und Begriffsfelder sowie kulturspezifische Konventionen des Kommunizierens verdeutlicht?
15. Werden intrakulturelle Unterschiede und transkulturelle Phänomene der Zielkultur(en) thematisiert?
16. Wird zum Kulturvergleich, insbesondere zum Vergleichen und Kontrastieren von Alltagssituationen, angeregt?

**D Zur Förderung der handlungsorientierten Dimension (*savoir faire*):**

17. Wird der Umgang mit nonverbalen Kommunikationsmitteln thematisiert und eingeübt?
18. Werden handlungsrelevante Verfahren des Rollenspiels zur Bewältigung interkultureller Begegnungssituationen vorgeschlagen?
19. Werden ganzheitliche Lernarrangements zum Eintauchen in den zielkulturellen Kontext (komplexe Lernaufgaben, Unterrichtsprojekte, simulation globale) vorgeschlagen?
20. Wird auf virtuelle Begegnungssituationen mit der Zielkultur hingewirkt?

Können einige der genannten Untersuchungsfragen für Lehrwerke auch auf Apps übertragen werden, so gestaltet sich dies für andere Schwerpunkte eher schwierig. Gibt es bei einem Lehrwerk in der Regel ein Inhaltsverzeichnis, so können Apps verschiedene Formen und Aufbauweisen haben, welche nicht zwingend ein transparentes Inhaltsverzeichnis aufweisen müssen, sodass es

bereits beim ersten Punkt zu Vergleichsschwierigkeiten kommen kann. Inwieweit eine App zum Einüben nonverbaler Kommunikationsmittel (Punkt 17) verhelfen kann, bleibt eher fraglich, genauso ist ein Unterrichtsprojekt mit zielkulturellem Kontext (Punkt 19) nicht möglich, da *M-Learning* gesondert vom Unterricht verwendet wird. So verlangen auch die Punkte 7 und 18 eine Thematisierung oder Einübung im Kurs – dies kann eine App nicht bieten. Auch Punkte 9 und 10 sind in einer App nicht zu erwarten, da die Bearbeitung bzw. Auswertung eher auf eine Diskussion/Reflexion im Kurs ausgelegt ist und eine App keine (offensichtliche) Möglichkeit zur Rückmeldung bietet. Auch der Zeit-Aspekt spielt bei vielen Punkten eine Rolle und ist für die vorliegende Arbeit erforderlich: Wie bereits beschrieben, ist ein zentrales Merkmal des *M-Learnings* auch der kurze Zeitrahmen einer Lektion. Hierbei stellt sich die Frage, ob in kurzen Einheiten ein Einbezug von beispielsweise (Lehrwerk-) Figures zur Identifikation möglich ist oder aber ob verschiedene authentische Materialien, z.B. literarische Texte oder Lieder (Punkt 11) herangezogen werden. Insgesamt ergibt die Recherche zu Sprachlern-Apps nur wenig Aufschluss darüber, welche Textformen in diesen verwendet werden. Scheint die Bearbeitung langer literarischer Texte im Kontext des Lernens in kleinen Einheiten als eher unwahrscheinlich, wird von den Anbietern selbst der Dialog in den Fokus gestellt. So wirbt *Babbel* mit „interaktiven Dialogen zu alltäglichen Situationen“ und *Busuu* verspricht Lernen „mit einer Kombination aus unseren AI-gestützten Ausspracheübungen und Dialogen, die von unseren Linguisten entwickelt wurden“.

Behandelt Usener (2016: 53) die Einübung dialogischer Sprachkompetenzen durch Rollenspiele, wird im Fragenkatalog nicht weiter auf die Bedeutung von schriftlich festgehaltenen Dialogen für die Ausbildung der interkulturellen Kompetenz eingegangen. Dialoge haben sich jedoch in den letzten Jahren zu einem vielversprechenden Forschungsgegenstand innerhalb der interkulturellen Kommunikation entwickelt, sodass diese im nächsten Abschnitt behandelt werden, um eine weitere Grundlage zur Evaluation von Sprachlern-Apps zu bieten.

## 2.3 Dialogisches Handlungsspiel

In den letzten Jahren hat sich der Fokus der Pragmatik von der Untersuchung von monologischen Sprechakten zur Forschung an Dialogen verschoben. So wird heutzutage angenommen, dass die kleinste kommunikative Einheit ein Handlungsspiel aus Aktion und Reaktion ist (vgl. Grein 2019: 130), was im *minimal action game* Modell nach Weigand (2010) diskutiert wird. Ausgangspunkt des Modells ist es nicht nur, die Aktion der/des Sprechenden und die Reaktion der/des Hörenden als Einheit zu betrachten, sondern auch den Menschen als Individuum mit all seinen Bedürfnissen, Wünschen und Fähigkeiten, Emotionen und sozialen Hintergründen zu sehen. Basierend auf dieser Ansicht ist davon auszugehen, dass keine Kommunikation ohne Grund zustande kommt, auch wenn diese nur bewirkt, dass Beziehungen durch beispielsweise Small-Talk gestärkt oder aufgebaut werden, denn „[s]peech, cognition, perception and having emotions interact“ (Grein 2018: 67).

Damit eine Konversation gelingen kann, muss der/die zweite Gesprächspartner/in also nicht nur die Äußerung der/des ersten mit einbeziehen: Hier

spielt neben der Situation (die auch wieder unterschiedlich wahrgenommen werden kann) auch die Persönlichkeit und die mit der Situation einhergehenden Emotionen eine Rolle. Menschen haben divergierende Einschätzungen und Wertungen von ein und derselben Situation (Grein 2019: 131).

Bezüglich dieser dialogischen Kommunikation bestehen Prinzipien, die offenbaren, wie das Handlungsspiel (*active and reactive speech act*) aufgebaut ist, jedoch keine festgelegten Regeln (vgl. ebd.). Diese Prinzipien sind von verschiedenen Faktoren geprägt und somit stark vom einzelnen Individuum abhängig, da hier „individual imprint“, „cultural imprint“, „perception & cognition“, „emotions“ und „principles of language use“ der Gesprächsteilnehmenden einfließen (Grein 2018: 86). Die Berücksichtigung der Gesprächsteilnehmenden in ihrer Gesamtheit als Mensch macht dieses holistisch-dialogische Modell aus (vgl. ebd.: 67). Prinzipien sind das Action Principle (AP), welches besagt, dass eine Äußerung einen kommunikativen Zweck beinhaltet und das

Dialogical Principle (DP) demzufolge Kommunikation auf dialogischen Aktionen beruht, welche aufeinander aufbauen (vgl. ebd.: 68). Dabei ist das „kommunikative Ziel eines jeden Handlungsspiels (oder auch jeden Dialoges) [...] zu einem gemeinsamen Verstehen und einer Verständigung (*understanding*)“ zu kommen (Grein 2019: 131). Dies erreichen Personen vielmehr durch deren kognitiven Hintergrund und deren Wissen, als durch explizite Aussagen, da „coherence is no longer in the text, it is given by the interlocutors“ (Weigand 2009: 160, zit. in Grein 2018: 68).

Wird der Mensch durch das Modell in seiner Ganzheit gesehen, ist seine kulturelle Prägung Teil dieser Ganzheit, sodass Action Games in verschiedenen Kulturen von verschiedenen Menschen äußerst unterschiedlich ausgegagt werden können. So kann es in (interkulturellen) Kommunikationssituationen zu Missverständnissen kommen, welche bei „kulturdivergierenden Sprechenden zunächst wahrscheinlicher als bei Sprecherinnen und Sprechern einer Kultur“ (Grein 2019: 131) sind. Diese Feststellungen spielen eine bedeutende Rolle für die interkulturelle Kompetenz und kommt Lernenden einer Fremdsprache zu Hilfe, da der authentische Spracherwerb und -gebrauch unter Beachtung der Forschungsergebnisse effektiv gefördert werden kann. Dies kann durch die Integration von *minimal action games* in Lehrwerken und durch Erklärung zur Unterschiedlichkeit von Äußerungen infolge soziokultureller Hintergründe in der Zielsprache erreicht werden (Grein 2018: 68, ebenso Grein 2019: 132).

Grein (2018) zeigt dies an modernen Lehrwerken für Deutsch als Fremdsprache auf und verweist auf die kulturell bedingte Andersartigkeit von Dialogen, welche schon bei kurzen Sequenzen festgestellt werden kann. Je nach kultureller Prägung können schon kurze Sprechakte, wie eine Begrüßung, bemerkliche Unterschiede aufweisen. Beginnt ein Italiener eine Unterhaltung mit einer flüchtig Bekannten mit „Ciao Bella“, kann diese Anrede „Hallo Schöne“ bei einer Deutschen bereits anzüglich wirken. Durch das Aufzeigen authentischer Dialoge, welche nicht nach Regeln, sondern Prinzipien aufgebaut werden, lernen Lernende hier schon in niedrigen Niveaustufen verschiedene Dialogmodelle kennen. So zeigt die Autorin, wie beispielsweise mehrere Optionen zur Vorstellung gegeben werden, die im Deutschen alle angemessen, direkt übersetzt in anderen Sprachen jedoch nicht unbedingt authentisch sind.



Die dargestellten Dialoge sind somit Beispiele für Kommunikationssituationen in der Fremdsprache – durch ihre Einbettung können Lernende jedoch noch mehr Wissen erwerben. So sind diese *minimal action games* häufig „supported by pictures or movie clips“ (Grein 2018: 72). Durch diese können Lernende auch nonverbale Signale beobachten und dabei „often indirectly, help learners to become aware of cultural differences, i.e. pictures demonstrating gender equality in Germany, the appropriate talking distance, eye contact, the display of emotions and frequency of touch“ (ebd.).

Ob und wie Dialoge nicht nur in Lehrwerken, sondern auch in Apps zur Ausbildung der interkulturellen Kompetenz beitragen und wie sich die in den letzten Abschnitten gesammelten Themenbereiche in einem Kriterienkatalog sammeln lassen, wird im nächsten Abschnitt behandelt.

### 3 Interkulturelle Kompetenz in Lern-Apps

Im vorherigen Abschnitt wurden Forschungsfragen aus der Lehrwerkanalyse wie auch der App-Evaluation behandelt, welche hier zusammengefasst Leitfragen für den praktischen Teil dieser Arbeit liefern. Ein vollständig ausformulierter Fragenkatalog wurde für diese Untersuchung als nicht sinnvoll betrachtet, da Lernapps sich in ihrer Form und im Aufbau stark voneinander unterscheiden können und deshalb, wie bereits erwähnt, nur bedingt vergleichbar sind (vgl. Godwin-Jones 2017: 6). Die folgenden Untersuchungsfelder können jedoch in Sprachlern-Apps erwartet und behandelt werden und bieten verschiedene Möglichkeiten, Lernenden interkulturelle Kompetenz nahe zu bringen. Die drei Apps *Babbel*, *Duolingo* und *Busuu* werden hierbei auf folgende Bestandteile untersucht:

- Landeskundliche Inhalte (Fachwissensvermittlung)<sup>2</sup>
- *Language variety* und intrakulturelle Unterschiede sowie DACH-Ansatz<sup>3</sup>

.....

2 Basierend auf Ciepielewska-Kaczmarek u. a. (2020: 77), Usener (2016: 90), Rosell-Aguilar (2017: 14).

3 Basierend auf Ciepielewska-Kaczmarek u. a. (2020: 77), Rosell-Aguilar (2017: 14).

- *Minimal action games* (authentische Dialoge)<sup>4</sup>
- Videos und Bilder (nonverbale und paraverbale Elemente)<sup>5</sup>

Können die hier genannten vier Themengebiete erwartet werden, werden weitere, wie etwa die von Usener (2016) genannten Punkte zur interkulturellen Kompetenz, ebenfalls beachtet, falls sich diese oder nur Ansätze dieser in den Apps finden lassen. Außerdem wird desgleichen in Betracht gezogen, inwieweit auf kulturelle Einheiten hingewiesen wird (vgl. Usener 2016: 53 und Ciepiewska-Kaczmarek et al. 2020: 77) bzw. ob diese in eine Lerneinheit integriert oder gesondert vermittelt werden.

Um die einzelnen Apps mit ihrer Ganzheit an Funktionen vergleichbar zu machen, wird zunächst der Start der App (also die ersten Schritte) sowie der Aufbau und das Inhaltsverzeichnis analysiert sowie Zusatzfunktionen, die außerhalb der abgegrenzten Aufgabenfelder verfügbar sind (Zusatzkurse und „Bonus Skills“). Um interkulturelle Inhalte innerhalb einer Lektion vergleichen zu können, wurde sich auf die ersten Lektionen der Apps beschränkt, welche dem Niveau A1 zuzuordnen sind. Um hier auch thematisch Vergleiche herstellen zu können, werden die Lektionen zum Thema „sich vorstellen“ bzw. „Essen/Essen gehen“ genauer untersucht, da diese Themenfelder frühzeitig vermittelt werden.

Die zu untersuchenden Sprachlern-Apps sind die mit am häufigsten heruntergeladenen Anwendungen (jeweils über 10 Millionen Downloads im *Google PlayStore*) und wurden von der Deutschen Gesellschaft von Verbraucherstudien mbH (DtGV) 2018 in der Studie „Sprachlernportale: Angebot & Preise, Lernspaß, Lernkomfort, Transparenz, Umfang der App“ mit weiteren vier Apps getestet. Hierbei wurde *Busuu* mit 85% Testsieger (2,0 | Gut), *Babbel* erreichte Platz 2 (84%, 2,1 | Gut) und *Duolingo* wurde Dritter (79%, 2,4 | Gut) (dtgv.de).

Die Installation und einige Lektionen sind kostenfrei, eine Premiumversion wird von allen drei Apps angeboten, auf die sich die folgende Untersuchung bezieht. Die Ausgangssprache, welche bei allen drei Anwendungen verwendet wurde, ist Englisch.

.....  
4 Basierend auf Grein (2017: 72, 2019), Usener (2016: 53).

5 Basierend auf Grein (2017: 72), Usener (2016: 53), Rosell-Aguilar (2017: 14).

### 3.1 *Babbel*

Gleich zum Start des Programms wird der/dem Lernenden die Frage gestellt, warum sie/er Deutsch lernen möchte; hier stehen die Unterpunkte „Travel“, „Career“, „Hobby“, „For friends and family“, „I'm planning to move“ oder „mental workout“ zur Auswahl. Die vorgeschlagenen Lektionen zu unterschiedlichen Niveau-Stufen zeigen einen lebensnahen Ansatz der App, klickt man beispielsweise auf „I'm planning to move“ schlägt die App Kurse auf B2 Niveau zu verschiedenen Dialekten vor.

Beginnt man mit der niedrigsten Niveaustufe, wird die erste Stufe „New-comer“ in verschiedene Kurse und „lessons“ aufgeteilt, die Titel wie z.B. „Hallo, wie geht's dir? Part 1“ tragen. Zunächst gibt es innerhalb des Inhaltsverzeichnisses keine Kennzeichnung interkultureller/landeskundlicher Themenbereiche. Gleich zu Beginn der ersten Lektion wird die Figur Freddy vorgestellt: Freddy „is an English journalist. He's just moved to Berlin. Join him and learn German!“ Mit diesem Charakter, mit dem die Lernenden sich identifizieren können, beginnt die erste Lektion, in der Begrüßung und Vorstellung eingeübt wird. Bereits in dieser ersten Lektion werden sprachliche Variationen angeboten und verschiedene Optionen zur Begrüßung aufgezeigt wie „Hi, Hey or Na?“ (*Babbel* Lektion 1, Hallo, wie geht's dir?). Die Lektion endet direkt mit einem ersten Dialog, in dem die ersten gelernten Sequenzen eingeübt werden. Hier wird die Rahmensituation beschrieben (Freddy zieht nach Berlin), sodass gleich der erste Dialog eine authentische Handlung mit begrenztem Wortschatz aufweist und somit ein Beispiel für ein erstes dialogisches Handlungsspiel stellt:

*Freddy is staying with a family friend until he finds his own place. He's introducing himself to his host...*

- ▶ *Hallo, ich bin Freddy*
- ▷ *Hallo Freddy. Ich bin Maria.*
- ▶ *Wie geht's dir?*
- ▷ *Mir geht's gut, danke. Und dir?*
- ▶ *Mir geht's auch gut.*
- ▷ *Willkommen in Berlin! Komm, ich zeige dir dein Zimmer.*
- ▶ *Danke.*

Werden die Dialoge selbst nicht mit Bildern unterstützt, wie beispielsweise in Lehrwerken (vgl. Grein 2018), so benutzt *Babbel* zur Einführung aller Vokabeln Fotos (Abb. 2), die Lernende nicht nur beim Einprägen der Vokabeln unterstützen, sondern auch einen ersten Eindruck in den deutschen Sprachraum vermitteln:



Abb. 2: Personalpronomen und langes „i“ (Original in Farbe)

Die abgebildeten Personen zeigen verschiedene Charaktere, aus Gruppen verschiedenen Alters, Genders und Ethnizität und vermitteln dadurch ein „buntes“ Deutschlandbild, welches verschiedene soziale Gefüge präsentiert und die Lernenden somit für intrakulturelle Unterschiede und Multikollektivität sensibilisieren kann. Diese Sensibilisierung durch Bilder findet auch dann statt, wenn andere Themenbereiche der Sprache im Fokus stehen. So wird in einem ersten Aussprachetraining das lange „i“ eingeübt, mit dem präsentierten Bild sehen die Lernenden jedoch auch, wie man sich im deutschsprachigen Raum im formellen Kontext begrüßt und werden hier für Distanzverständnis und Körperhaltung sensibilisiert.

Diese visuelle Einführung in kulturelle Begebenheiten setzt *Babbel* fort, indem nicht nur Bilder, sondern zusätzlich auch Videos angeboten werden (siehe Abb. 3). Gleich in der ersten Lern-Stufe werden in der Web-Version insgesamt drei Video-Lektionen als Zusatzmaterial angeboten. Hier wird ein weiterer Charakter eingeführt: Julia aus Minnesota, die ebenfalls nach Berlin zieht. In den amüsant gestalteten Videos werden Alltagssituationen gezeigt, in denen gelernte Vokabeln, Grammatikregeln und ähnliches wiederholt werden. Durch Julia gibt es hier somit auch einen weiteren Charakter, mit dem sich die Lernenden identifizieren können. Im Gegensatz zu Freddy ist Julia weiblich und Amerikanerin, wodurch weitere Lernendengruppen eine Bezugsperson finden.

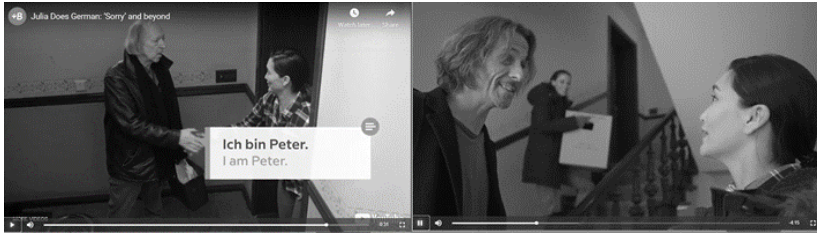


Abb. 3: Video „Sorry and beyond“ (Original in Farbe)

Die gedrehten Szenen (auf der Straße, im Treppenhaus) erlauben Lernenden Rückschlüsse auf das Leben in Berlin und auch hier können kulturelle Eigenschaften wie Distanzverhalten oder Mimik und Gestik beobachtet werden. Auch die hier gezeigten Dialoge zwischen deutschen Muttersprachler/innen untereinander oder mit Julia stellen authentische und lebensnahe Dialogsituationen dar.

Des Weiteren erlauben diese Szenen auch eine anschauliche Vermittlung von Fachwissen bzw. landeskundlichen Inhalten. Hier wird zum Beispiel die „Sonntagsruhe“ thematisiert, sodass kulturspezifische Phänomene und Begriffsfelder angesprochen werden (vgl. Usener 2016). Julias Nachbar klopft an der Tür und beschwert sich über den Krach beim Einzug am Sonntag, bevor er wieder geht. Julias deutsche Freundin erklärt ihr diesen Critical Incident und sie bereiten zusammen eine Entschuldigung vor, sodass hier die sprachlichen mit den landeskundlichen Besonderheiten verbunden werden und den Lernenden hierdurch alltagsnahe Situationen nähergebracht werden.

Der Aufbau aus Wortschatzeinführung, Einüben und alltagsnahen Dialogen bleibt bei allen Themen erhalten, sodass neu eingeführte Worte/Sequenzen zunächst einzeln präsentiert werden und dann im Kontext verwendet werden. Auch beim zweiten untersuchten Thema „Essen und Trinken“ bzw. „Restaurantbesuch“ sind die Dialoge authentisch und abwechslungsreich konzipiert und in den Videosequenzen werden verschiedene Szenarien eines Dialogs in einem Café behandelt, sodass Lernende mit einer Auswahl an Sprachmitteln vertraut gemacht werden, also authentische *minimal action games* präsentiert bekommen: So werden hier „Soll's noch was dazu sein?“, „Sonst noch was?“

und „Ist das alles?“ als Fragen eines/einer Bäckereiverkäufers/-verkäuferin eingeführt.

In der Lektion selbst erfolgen außerdem Erklärungen zu landeskundlichen Themen, in diesem Fall beispielsweise zum Trinkgeld. Hierbei wird die landeskundliche Fachwissensvermittlung mit den neu erlernten sprachlichen Mitteln verknüpft:

### **Paying and tipping**

*In Germany it is very common to split the bill. You'll usually be asked:*

*Zusammen oder getrennt? Together or separate?*

*If you want to pay separately, you answer:*

*Getrennt bitte! Separate please!*

*It's common to leave a tip of around 10%. If you wanna tell the server to keep the change, you simply say*

*Stimmt so. Keep the change (lit. It's correct like this).*

*(Babbel: Lesson 9: Was möchten Sie? Part 2 Paying and tipping)*

Verweist der Titel „Paying and tipping“ zwar auf eine alltägliche Situation, ist diese kurze Einheit aber nicht explizit als interkultureller oder landeskundlicher Einschub gekennzeichnet. Viel mehr fließen diese Erklärungen – sei es schriftlich festgehalten oder im Video besprochen – in die einzelnen Lektionen mit ein, so auch in der Videolektion, in der thematisiert wird, dass Cafés und kleinere Läden nicht unbedingt Kreditkarten akzeptieren, sondern hier bar bezahlt wird.

Auch regionale Besonderheiten lassen sich innerhalb der Lektionen finden. Behandeln die Lernenden zum Beispiel weiter das Thema „Essen“, werden bei einem Restaurantbesuch in Stuttgart regionale Spezialitäten wie die schwäbischen Käsespätzle thematisiert. Mit Bildern und landestypischen Gerichten wird den Lernenden die „deutsche Küche“ nähergebracht. Das Thema „Essen“ bleibt jedoch nicht auf die deutsche Küche beschränkt: Durch Beispiele wie „Sushi essen gehen“ präsentiert die App auch andere Küchen, die durchaus im deutschsprachigen Raum beliebt sind. Landeskundliche/interkulturelle Themen werden hierbei implizit und integrativ behandelt, also mit dem Thema der Einheit verknüpft.

Ferner findet sich unter der Kategorie „more courses“ eine Einheit zu Ländern und Traditionen, die explizit als solche bezeichnet wird: „Countries and Traditions. In these courses, you won't just learn the language. You'll also gain useful knowledge about Germany.“ Innerhalb dieser Kategorie können Lernende beispielsweise Dialekte des Deutschen kennenlernen: Hierbei gibt es eine Auswahl aus Berlinerisch, Hessisch, Oberfränkisch, Schwäbisch, Sächsisch, Plattdeutsch und Bayerisch. Es wird erklärt „The Bavarian dialect is spoken in Southern Germany, in the area around Munich and also in some parts of Austria“, weitere DACH-relevante Dialekte werden jedoch nicht vermittelt. Innerhalb der verschiedenen Sonder-Lektionen wird nicht nur auf dialektische Bezeichnungen eingegangen, sondern auch landeskundliches Fachwissen vermittelt. Als Beispiel erfolgt die Vermittlung und Erklärung des „Spätis“ in Berlin:

#### **No Berlin without a Späti**

*Often, colloquial expressions are of regional origin, such as the Späti which is actually short for Spätkauf and literally means late-purchase.*

*As Spätis originated in the former GDR to supply shift workers, you find them mostly in Berlin and east Germany. Fun fact: Legally, shops in Germany are not allowed to sell goods on Sundays. So, technically, Spätis are violating the law.*

► *Lass uns mal zum Späti gehen und noch ein Bier holen!*

*(Babel: No Berlin without a Späti)*

Auch diese regionalen Besonderheiten werden in Dialogen verfestigt. Besonders bei den Kursen in dieser Kategorie wird auch die Jugendsprache behandelt – hierzu existiert auch ein weiterer Kurs der Kategorie „Specials“: „Talk like a Babbo“. In Geschichten, bestehend aus alltagsnahen Situationen, kommen hier auch intrakulturelle Unterschiede zum Vorschein, beispielsweise wird die Gay- und Technoszene mit in die Handlung integriert, indem sie Schauplatz eines Dialogs ist („Berghain ist he most famous (gay) club in Berlin – and probably Europe. Some describe it as the *world capital of techno*.“) und auch die gezeigten Bilder zu Erklärungen stellen ein vielseitiges Deutschlandbild dar:

Fun fact: *Kopfkino* can also lead to rather horrible scenarios playing in your head. Sorry.

Close

### What if...?

*Kopfkino* implies a little bit more than just *imagination*...

So, let's say you're at your local *Späti*, when you notice the person in front of you. They catch your eye and your *Kopfkino* starts *rolling*:

You can see your hands touching when you both go to reach for the same beer bottle.

In the next scene, you're on a date together – laughing, kissing, and then the camera cuts to... your wedding!

Wenn ich Mina mit dem Typ sehe, springt bei mir gleich das *Kopfkino* an.

When I see Mina with that guy, my imagination immediately starts running wild.



Abb. 4: *Kopfkino* (Original in Farbe)

Die App bemüht sich hier um eine lebhaft und in Deutschland realitätsnahe Vermittlung der sprachlichen Inhalte, welche Raum für interkulturelle Rückschlüsse geben. So wird in dieser kurzen Sequenz (Abb. 4) das Kennenlernen zweier Personen, wie es sich in Deutschland abspielen könnte, Alkoholkonsum und durch das Bild eine offene Einstellung zu Homosexualität abgebildet – Themen, die in anderen Kulturen auf andere Weise behandelt werden könnten. Bemerkenswert ist hierbei vor allem, dass besonders das Thema Homosexualität in Lehrwerken des DaF-Unterrichts eher selten aufgegriffen wird. Lediglich in Begleitmaterialien für den Deutsch als Zweitsprache-Unterricht (z.B. *Mein Leben in Deutschland* oder *Vielfalt leben*) wird diese Thematik aufgegriffen und meistens im Zusammenhang mit dem Thema Lebensformen oder Diskriminierung thematisiert. Die App hingegen lässt das Thema beiläufig in die Handlung einfließen, die Nennung des Begriffs „(gay) club“ und das Bild zweier tanzender Männer wirken durch ihre Beiläufigkeit eher normal und werden nicht als Tabubruch thematisiert. Diese Inhalte werden von der App implizit vermittelt, eine Reflexion/Evaluation bleibt allerdings aus, sodass Lernende die vermittelten Inhalte allein verarbeiten müssen.



### 3.2 *Busuu*

Zu Beginn der App hat der/die Lernende die Auswahl zwischen „Complete German“ oder „German for Travel“. Wählt man die erste Lektion von „Complete German“, beginnt der Kurs für Lernende ohne Vorkenntnisse und wird mit den Flaggen Deutschlands, Österreichs und der Schweiz eingeleitet. Der von der App aufgezeigte DACH-Ansatz wird auch gleich in der ersten Lektion umgesetzt: So werden die Flaggen Deutschlands, Österreichs und der Schweiz gezeigt und die App gibt an: „Today we'll learn about... 1. **Formal** titles and greetings 2. **Regional** greetings in [Flaggen Deutschlands, Österreichs und der Schweiz]“ (*Welcome to Busuu's German course*).

So werden gleich zu Beginn des Kurses verschiedene intralinguale Unterschiede hervorgehoben und thematisiert. Die Begrüßungen „Moin“, „Grüß Gott“, „Servus“ und „Gruetzi“ werden vorgestellt und die verschiedenen Regionen benannt, in denen diese Variationen verbreitet sind, dabei werden fotografische Abbildungen (Abb. 5) eingesetzt:

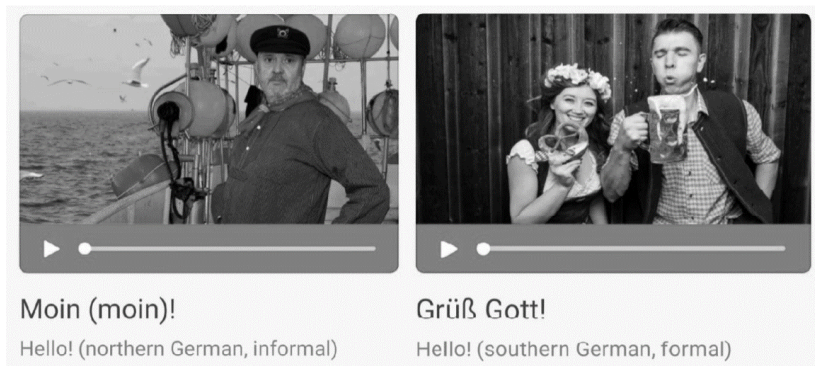


Abb. 5: Begrüßungen (Original in Farbe)

Sind die gezeigten Fotografien für diese Begrüßungsarten stereotypisch, zeigen andere Abbildungen zu Begrüßung und Verabschiedung ein vielseitiges Bild deutschsprachiger Länder mit Menschen aus verschiedenen Kollektiven. Diese werden auch im zur Lektion gehörigen Video betont. Innerhalb der Lektion wird auf den *Youtube* Channel von *Busuu* und das Video „12 German greetings

they WON'T teach you in school“<sup>6</sup> verwiesen. In diesem Video erklärt „German Language Expert“ Anna verschiedene Formen der Begrüßung und Verabschiedung in unterschiedlichen Regionen, auch unter Beachtung Österreichs und der Schweiz. So wird hier ebenfalls auf umgangssprachliche Ausdrücke wie „Na?“ oder Jugendsprache wie „Was geht ab?“ eingegangen. Frontal, aber auf amüsante Weise, werden diese Formen eingeführt und in Beispielen verwendet. Die Erklärungen finden hierbei auf Englisch statt. Dadurch, dass nur eine Person im Video präsent ist, können Interaktion zwischen Personen und beispielsweise Distanzen, deren Beobachtung Lernenden zu einer Sensibilisierung verhelfen können, nicht gezeigt werden, jedoch geht die Sprachexpertin auch auf Betonung und Mimik ein.

Zurück in der App unterstützen Bilder (Abb. 6) das Verständnis vom Miteinander im deutschsprachigen Raum:

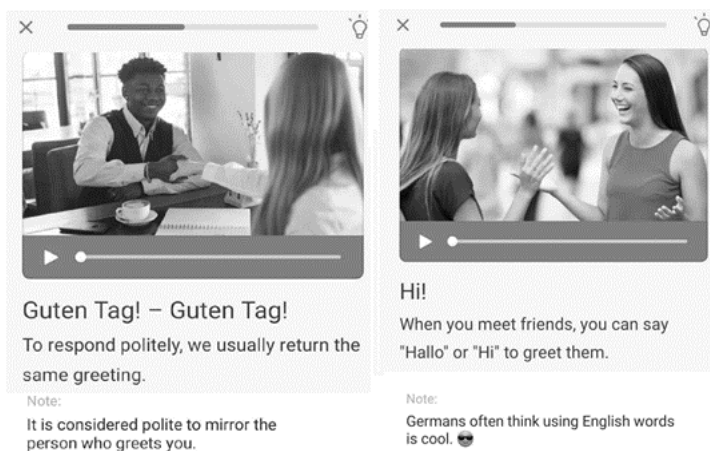


Abb. 6: Begrüßungen und Tipps (Original in Farbe)

Hinweise und Erklärungen zu interkulturellen Handlungsweisen und landeskundlichen Themen finden sich innerhalb der Lektionen und sind in die jeweiligen Themenbereiche integriert und implizit. Hierbei nutzt die App einen spielerischen, informellen Umgang, um landeskundliche und interkulturelle

6 Video abrufbar unter: <https://www.youtube.com/watch?v=flfayo4HiJ0>, 21.10.2020.

Inhalte zu adressieren (Abb. 6) und kulturspezifische Konventionen des Kommunizierens zu verdeutlichen. Auch ermutigt *Busuu* Lernende, landestypische Dinge zu probieren, wie zum Beispiel die Apfelsaftschorle im folgenden Beispiel (Abb. 7). Durch Appelle wie diese „go ahead and try one of the many types of *Schorle*!“ werden Lernende dazu ermutigt, sich auf die neue Kultur und ihre Besonderheiten einzulassen. So werden auch in den Lektionen zum Bar- bzw. Restaurantbesuch weitere landeskundliche Erklärungen angeboten und das Verständnis dieser hier auch überprüft:

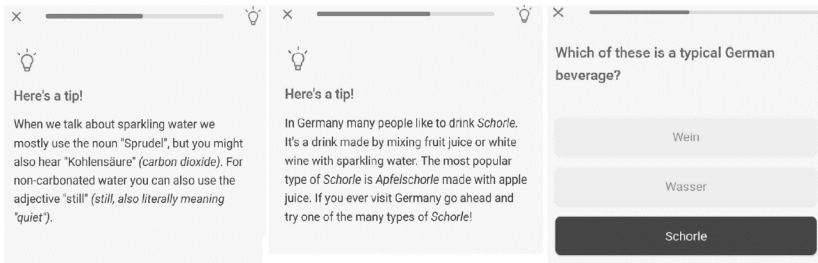


Abb. 7: Sprudel und Schorle (Original in Farbe)

Leistet die App zwar keine Selbstevaluation zur Einstellung der Lernenden oder Reflexion der vermittelten interkulturellen Inhalte, werden hier jedoch nicht nur die sprachlichen Inhalte, sondern auch das vermittelte landeskundliche Wissen abgefragt.

Die Dialoge behandeln thematisch und den Sprachmitteln angemessen den Inhalt der Lektion und finden in einem authentischen Rahmen statt:

*Bea: Dieser Wein ist sehr gut!*

*Martin: Soll ich dir eine Flasche bestellen? Ich trinke lieber Bier.*

*Bea: Danke! Ein Glas ist genug. Wollen wir auch Wasser bestellen?*

*Martin: Ja klar, mit oder ohne Sprudel?*

*Bea: Mit, danke.*

*Martin: Gut, dann nehmen wir eine Flasche Sprudelwasser mit zwei Gläsern, ein Glas Wein und ein großes Bier.*

*(Busuu Lektion 44: Shall we sit at a bar: Dialogue)*

Wird in der englischen Einleitung die Situation beschrieben („Bea and Martin are at a bar talking about the drinks they want to order“), ist der Dialog, den die Lernenden hören und lesen können, passend zum lockeren Setting und authentisch vorgetragen. Die erarbeiteten landeskundlichen Informationen zum Sprudelwasser werden auch hier thematisiert, sodass die Lernenden sich weiter mit diesem Thema auseinandersetzen können. Die Charaktere in den Dialogen wurden hierbei von der App nicht eingeführt, die verwendeten Bilder entsprechen zwar dem Gender der Personen, stellen sie jedoch nicht in der beschriebenen Situation dar, was eine Identifikation mit den Charakteren erschwert.

Im Zusatzkurs „German for travel“ werden Lektionen zu verschiedenen Schwerpunkten wie Transport, Restaurantsuche, Organisation eines Ausflugs und Sightseeing angeboten. Auch hier werden Dialoge in möglichen Situationen während eines Besuchs in deutschsprachigen Ländern präsentiert und oft verschiedene Optionen gegeben, was ein Einheimischer/eine Einheimische antworten könnte. Fragemöglichkeiten sind zum Beispiel: „Gibt es ein gutes Restaurant hier in der Nähe?“ und „Können Sie mir ein Restaurant hier in der Nähe empfehlen?“ Nach der Einführung dieser sprachlichen Mittel werden mögliche Antworten präsentiert: „Sie könnten es mit dem Café Havanna versuchen.“, „Es heißt Café Havanna“ oder „Es gibt ein tolles Restaurant in der Grünstraße“. Hiermit lernen Lernende verschiedene mögliche Antworten kennen, die zu einem authentischen *minimal action game* verwendet werden können und somit für den Gebrauch dieser Mittel sensibilisiert werden.

Zusätzlich zu den Kursen bietet *Busuu* eine weitere in die App integrierte Funktion an: „Social“. Hier können Lernende durch Aufgaben Kontakt zu anderen Lernenden aufnehmen, die die gewählte Fremdsprache ebenfalls lernen. Dies dient laut App zum einen als Unterstützung zum Erlernen der Sprache, deutschsprachige Benutzer/innen haben jedoch zum anderen die Möglichkeit, Aufgaben zu bewerten und zu korrigieren. Hierbei gibt es bei einigen Lektionen die Aufgabenstellung „Conversations“. In dieser haben Lernende die Auswahl, die Aufgabe schriftlich oder mündlich abzulegen und zu einer bestimmten Situation/Fragestellung eine Äußerung zu verfassen. Die Aufgabe, entweder schriftlich festgehalten oder mit dem Mikrophon aufgenommen, wird dann zur Korrektur an die „Community“ gesendet. Hier können Lernende,

die angegeben haben, dass sie Deutsch sprechen, die Aufgabe mit bis zu fünf Sternen, Kommentaren und Verbesserungen korrigieren und ein Feedback geben. Es gibt hier somit einen direkten Austausch mit anderen Sprecher/innen der Zielsprache, teils Muttersprachler/innen, sodass ein interkultureller Austausch auf diese Weise stattfinden kann. Die Funktion ermöglicht somit eine Kontaktaufnahme, die zum weiteren interkulturellen Austausch anregen kann. Hierbei soll jedoch angemerkt werden, dass die Korrekturen unkontrolliert an die Lernenden weitergegeben werden, hier also Fehler innerhalb des Feedbacks nicht ausgeschlossen werden können.

In Bezug auf die interkulturelle Kompetenz kann diese Option jedoch als Ansporn für den Austausch mit der Zielkultur gesehen werden und die Ausprägung dieser somit fördern.

### 3.3 *Duolingo*

Startet man die App *Duolingo* und gibt an, dass man noch keine Deutsch-Kenntnisse hat, werden verschiedene Kategorien aufgezeigt. Diese heißen „Basics 1“, „Intro“ oder „Phrases“ und können durch Abschluss der vorherigen Lektionen freigeschaltet werden. Die einzelnen Lektionen sind so aufgebaut, dass zunächst Wortpaare in Deutsch und Englisch gefunden werden müssen, danach werden Wörter und Phrasen geübt. Bei der Lektion „Intro“, also der Vorstellung, fällt auf, dass zwar verschiedene Möglichkeiten zur Vorstellung der eigenen Person gegeben werden („Ich bin Hans“, „Hallo, ich heiße Julia“ vgl. Intro), hier findet aber kein Dialog und somit kein *minimal action game* statt. Auch in den Lektionen zum Thema „Essen“ (Food 1, Food 2) werden keine Dialoge aufgeführt. Zwar werden Sätze zum Themenfeld vorgestellt und durch die Übungen (Übersetzen, teils durch die Anordnung durch Wörter) eingeübt, jedoch bleiben auch hier dialogische Handlungsspiele komplett aus. Phrasen wie „Die Speisekarte ist schmutzig“ oder „Ich koche nicht ohne Wasser“ werden ohne Zusammenhang präsentiert. Jedoch bietet *Duolingo* eine weitere Kategorie „DuolingoStories“ (Abb. 8). Mit diesen Kurzgeschichten können Lernende, laut App, Lese- und Hörverstehen verbessern. Die Geschichten bestehen aus Dialogen, die durch kurze Verständnisfragen unterbrochen werden:

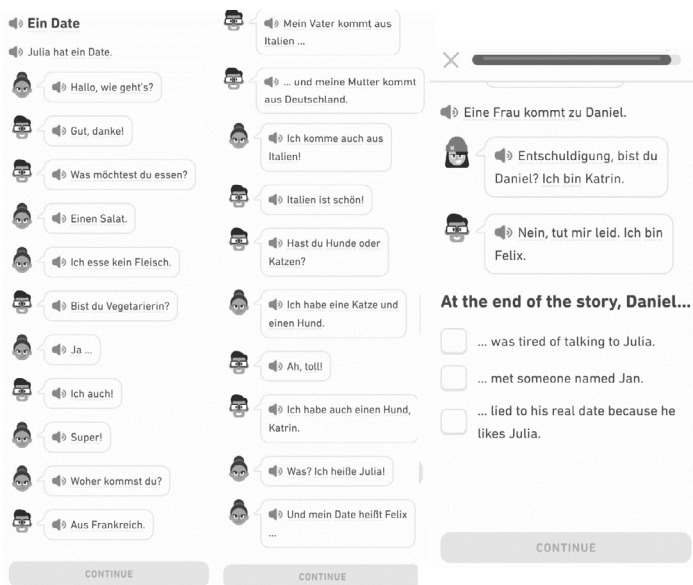


Abb. 8: *Duolingo Stories: A Date* (Original in Farbe)

So werden hier Dialoge angeboten, die in authentischen Situationen im deutschsprachigen Raum zu finden sind (ein Blinddate wie im Beispiel dargestellt ist in Deutschland nicht ungewöhnlich), die Authentizität der Dialoge selbst ist jedoch fraglich. Hier geht es in erster Linie darum, Sprachmittel zu benutzen und zu wiederholen. Dies leistet die App in einer amüsanten Weise, interkulturelle Themenfelder werden hier jedoch nicht behandelt.

Die in dieser Kategorie, aber auch in den Lektionen gezeigten Grafiken sind immer Illustrationen bzw. Animationen, keine authentischen Fotografien (Abb. 9). Hierbei fällt auf, dass sich die Bilder wiederholen und nicht unbedingt zu den Aussagen passen, sodass Lernende hierbei keine Rückschlüsse auf kulturelle Themen ziehen können:

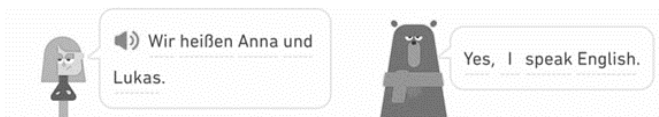


Abb. 9: Illustrationen (Original in Farbe)

*Duolingo* begründet die Auswahl „unerwartete[r] Inhalte“ (*duolingo.de*) damit, dass diese Lernende veranlassen, „sorgfältig über die Sprache nachzudenken, die sie lernen“ (ebd.). Zu diesen unerwarteten Inhalten kann auch die Verwendung einer Bären-Illustration gewertet werden. Dies trägt in diesem Falle jedoch nicht zur Sensibilisierung durch Bilder bei. Auch unterstützende Videos werden in der App nicht verwendet und Zusatzmaterialien wie Podcasts gibt es zwar für Spanisch oder Französisch, jedoch nicht für Deutsch (vgl. <https://podcast.duolingo.com/>). Bleibt die Vermittlung interkultureller Inhalte in den einzelnen Lektionen somit weitestgehend aus, kann vor jeder Lektion jedoch ein „Tipp“ abgerufen werden, der Ratschläge für die kommende Lektion beinhaltet und bei manchen Lektionen auch einige landeskundliche Inhalte vermittelt, aber kein differenziertes Bild liefern, wie in der folgenden Abbildung zum Thema Essen zu sehen ist:

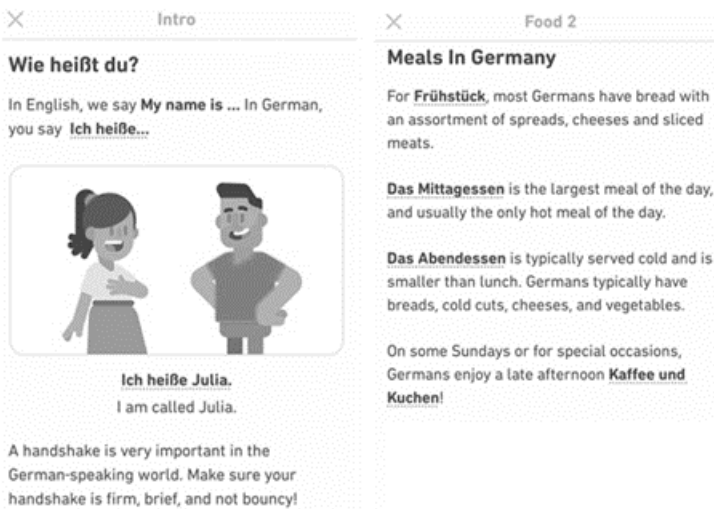


Abb. 10: Tipps (Original in Farbe)

Diese Einleitungen in die Lerneinheiten sind nicht explizit als solche ausgewiesen und durch die Voranstellung nicht Teil der eigentlichen Lektion, sodass keine der von Ciepielewska-Kaczmarek et al. (2020: 77) beschriebenen Kategorien gänzlich zur Einbettung der landeskundlichen Einheit passt. Da die Erklärun-

gen jedoch die Thematik der Lektion aufgreifen, kann man hier bestenfalls von einer vorangestellten, additiven Erklärung sprechen. In den untersuchten Lektionen wurden somit allgemeine Informationen zum deutschsprachigen Raum gegeben, der DACH(L)-Ansatz wird (in den untersuchten Modulen) hierbei jedoch nicht verfolgt. Die hier gezeigten Illustrationen können als Karikaturen in alltäglichen Situationen gesehen werden, unterstützen die Lernenden aber dadurch nur sehr bedingt in der Sensibilisierung zu interkulturellen Themen wie spezifische Mimik, Gestik oder auch Distanzverhalten.

*Duolingo* bietet desweiteren „Bonus Skills“ an, die für erspielte „Lingots“ freigeschaltet werden können. Hierbei können Lernende zwischen „Idioms and Proverbs: Learn some German idioms and proverbs“ und „Flirting: Do you believe in love at first sight? Learn to flirt in German“ wählen.

Bei der Untersuchung der Themenbereiche fällt auf, dass auch hier keine Dialoge, sondern nur einzelne, kontextlose Phrasen, wie „Der nächste Kaffee geht auf mich“, eingeübt werden. Einige dieser Redemittel werden im Alltag zwar verwendet, jedoch bekommen Lernende keinerlei Informationen darüber, wie sie zu benutzen bzw. wann diese angebracht sind. Aussagen wie „Du siehst aus wie meine nächste Freundin“ können schnell anzüglich oder unpassend wirken. Dadurch, dass Lernende nicht zur Reflexion angeregt oder durch Erklärungen für die Anwendung der Redemittel sensibilisiert werden, bleibt nicht nur der Zuwachs an interkultureller Kompetenz aus, auch der Erfolg bei Einsatz der Sätze in realen Situation bleibt fraglich.

## 4 Zusammenfassung der Ergebnisse

Die vorangegangene Untersuchung der drei Sprachlern-Apps hat Aufschluss über verschiedene Themenbereiche gegeben, die in diesem Abschnitt vorgestellt werden. So konnte gezeigt werden, dass Apps sich stark in Aufbau und Form voneinander unterscheiden, was einen direkten Vergleich dieser schwierig gestaltet. Besonders durch verschiedene Kursarten (z.B. Deutsch für den Urlaub, Zusatzkurse) und verschiedene Zusatzangebote wie *Youtube* Channels, Webansichten, Chats und Podcasts, gestaltet sich ein direkter Vergleich kompliziert.



Auch in Bezug auf die Vermittlung der interkulturellen Kompetenz gab es somit erhebliche Unterschiede in den untersuchten Teilen der drei ausgewählten Apps, welche in der folgenden Tabelle zusammengefasst werden:

	<i>Babbel</i>	<i>Busuu</i>	<i>Duolingo</i>
Landeskundliche Inhalte (Fachwissensvermittlung)	<input checked="" type="checkbox"/> Durch Einschübe in Lektionen integriert und implizit	<input checked="" type="checkbox"/> Werden teilweise zusätzlich in den Lektionen abgefragt	/ Teilweise als vorangestellte Erklärungen
Language Variety und intra-kulturelle Unterschiede so wie DACH-Ansatz	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Minimal Action Games (authentische Dialoge)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Videos und Bilder (non-verbale und para-verbale Elemente)	<input checked="" type="checkbox"/> Fotos in der App; Videos als Zusatz- Lektionen	<input checked="" type="checkbox"/> Fotos in der App; Auf <i>YouTube</i> Videos wird verwiesen	<input checked="" type="checkbox"/>

Während sich *Duolingo* stark auf die Vermittlung von Vokabular beschränkt und nur wenige landeskundliche Inhalte einen Teil der Fachwissensvermittlung darstellen, gehen die Anwendungen *Busuu* und *Babbel* durch verschiedene Mittel über diesen ersten Schritt der Wissensvermittlung hinaus und können Lernende bei der Sensibilisierung für interkulturelle Schwerpunkte unterstützen.

Durch landeskundliche Erklärungen und Themenfelder, die bei *Babbel* und *Busuu* in die Lektion integriert und bei *Duolingo* vorangestellt sind, geben die Anwendungen Informationen zu kulturellen Phänomenen und Besonderheiten. Besonders ins Detail gehen diese Erklärungen bei *Babbel*, da hier u.a. alltagsnahes Fachwissen (Bargeldzahlungen, Trinkgeld) behandelt werden. *Busuu* fragt die vermittelten Informationen direkt ab und sichert hier das Verständnis dieser Themenbereiche (vgl. das Thema Saftschorle).

Im Punkt Language Variety zeigt *Duolingo* in den untersuchten Lektionen im Gegensatz zu den anderen beiden Apps keine DACH-Ansätze. Diese werden

von *Babbel* in gesonderten Zusatzkursen hervorgehoben – wobei der Fokus dennoch eher auf deutschen Varietäten liegt – während *Busuu* sie in die Lektionen mit einbindet und auch im zusätzlichen Video thematisiert. Beide Apps leisten hier positive Ansätze zur Sensibilisierung der Lernenden.

Im Bezug zu den präsentierten Dialogen fällt bei *Duolingo* auf, dass die App als einzige keine Dialoge innerhalb der einzelnen Lektionen bearbeitet. Werden Dialoge in Kurzgeschichten verwendet, sind diese weniger authentisch als die in die Lektionen integrierten Dialoge bei *Babbel* und *Busuu*. Besonders *Babbel* fällt hier positiv auf, da es Dialoge in einen authentischen Rahmen setzt und mit den lektionsbegleitenden Charakteren Identifikationsmöglichkeiten schafft. Lernende bekommen hier ein alltagsnahes Deutschlandbild aufgezeigt, da in den Dialogen authentische Sprache benutzt und kulturspezifische Informationen aufgegriffen werden, sodass *Babbel* zum Ausbau der interkulturellen Kompetenz beiträgt.

Durch Bilder (bei *Busuu* und *Babbel*) zeigen die Anwendungen alltägliche Situationen und präsentieren ein vielseitiges Deutschlandbild – wobei hier bei einigen Bildern auf die Vermeidung von Stereotypen zu achten ist (*Busuu*). Besonders *Babbel* ist durch seine realitätsnahen Fotos hervorzuheben. Tabuthemen wie Homosexualität werden durch die Verwendung von Fotos vermittelt, jedoch nicht als solche thematisiert, sodass Lernende diese nicht als Tabubruch wahrnehmen. *Duolingo* ist die einzige App ohne Fotos und benutzt Bilder, die zufällig gewählt scheinen und den Lernenden keine Aufschlüsse zur Kultur in deutschsprachigen Ländern geben.

So präsentieren *Busuu* und vor allem *Babbel* situationsnahe Dialoge und gehen mit zusätzlichen Video-Lektionen, in denen spezifische Themen oder aber auch Schwierigkeiten (critical incidents) angesprochen werden, noch über dies hinaus (*Babbel*). Hierbei gibt es keine konkreten Fragestellungen oder Aufgaben zum Vollzug eines Perspektivwechsels oder „zum Vergleichen und Kontrastieren von Alltagssituationen“ wie Usener (2016: 53) bei Lehrwerken untersucht. Die Selbstevaluation wird von den Lernenden in keiner der Apps eingefordert, sodass interkulturelle Themen zwar behandelt werden, Lernende mit der Be- und Verarbeitung dieser jedoch allein gelassen sind. Diese unterstützte Reflexion, welche in Apps nicht gegeben ist, kann im DaF/DaZ-Unterricht durch die Lehrkraft oder Mitlernende angeregt und

eingeleitet werden. Der Austausch über Andersartigkeiten, die Besprechung verschiedener Phänomene sowie die Möglichkeit, Fragen über die Zielkultur (und auch die eigene) zu stellen, können in Gesprächen miteinander, nicht aber alleine über eine digitale Lernanwendung geklärt werden, da der Austausch, beispielsweise über *Busuu*'s „social“ sehr beschränkt bleibt. Auch die allgemeine Behandlung des Konzepts „Kultur und Kommunikation“, die für eine Ausprägung der interkulturellen Kompetenz notwendig ist, leisten die untersuchten Anwendungen nicht, sodass ein wichtiger Punkt, die Erkenntnis des eigenen Ethnozentrismus, nicht thematisiert wird.

## 5 Fazit

Hat die Untersuchung der verschiedenen Apps teilweise Ansätze zur Vermittlung interkultureller Kompetenz gezeigt, konnte jedoch auch festgestellt werden, dass eine Reflexion des Präsentierten ausbleibt. Sprachlern-Apps, die interkulturelle Kompetenz aufgreifen, können somit nicht als uneingeschränkter Ersatz des DaF/DaZ-Unterrichts gelten. Jedoch können sie diesen sehr wohl unterstützen. Hierfür plädiert auch Godwin-Jones (2017: 13):

Important in that process is presenting persuasive illustrations of learning connected to students' lives (present and future) and to bring those experiences into the classroom. The most effective way to do that may be through the smartphone they likely all own. The opportunity is to leverage those digital devices and online experiences to enable and encourage in our student's life-long learning, learner autonomy, and critical digital literacy.

Bekommen DaF/DaZ-Lernende durch die Nutzung von Apps neue Anregungen zu interkulturellen Themen, welche danach mit Mitlernenden und Lehrkräften thematisiert und reflektiert bearbeitet werden können, sind bestimmte Sprachlern-Apps wie *Babbel* und *Busuu* nicht nur in anderen Feldern des Spracherwerbs ein unterstützender Faktor beim Fremdspracherwerb, sondern auch im Erlangen und Ausbau der interkulturellen Kompetenz. Besonders in

Zeiten, in denen kein Präsenzunterricht geleistet werden kann, können diese Apps kurzfristige Alternativen bieten.

Durch die fortwährende Weiterentwicklung von digitalen Anwendungen bleibt jedoch ein großes Forschungsgebiet erhalten, welches Raum für weitere Analysen lässt. So können weitere Untersuchungen klären, ob sich die verschiedenen Apps in der Vermittlung der interkulturellen Kompetenz weiterentwickeln bzw. weiterentwickelt haben, so wie es bei gedruckten Lehrwerken bereits der Fall war. Ein weiteres interessantes Forschungsthema ist außerdem die Frage, ob Nutzer/innen der Apps *Busuu* und *Babbel* die aufgezeigten Mittel als interkulturelle Sensibilisierung wahrnehmen. Da ein Zuwachs an digitalen Lernanwendungen auch in der Zukunft als wahrscheinlich gilt, gibt es im Feld der Vermittlung interkultureller Kompetenz also nicht nur weitere Forschungsthemen für Wissenschaftler/innen – viel mehr besteht hier auch für die Verantwortlichen dieser Apps noch viel Potential zum Ausbau dieser unabdingbaren Komponente zum Fremdspracherwerb.

## Bibliografie

- BIECHELE, B. (2006): „Anmerkungen zum interkulturellen Bildverstehen“, in: *Intercultural Journal* 5/1: 17–50.
- CATALDO, C. (2017): *Die Vermittlung und Entwicklung interkultureller Kompetenzen im DaF-Unterricht*, abrufbar unter: <https://core.ac.uk/download/pdf/141499967.pdf>, 30.07.20.
- CHEN, X. (2016): „Evaluating Language-learning Mobile Apps for Second-language Learners“, in: *Journal of Educational Technology Development and Exchange (JETDE)* Vol. 9/2: 39–51, abrufbar unter: <https://aquila.usm.edu/jetde/vol9/iss2/3>, 30.07.20.
- CIEPIELEWSKA-KACZMAREK, L., JENTGES, S. & TAMMENGA-HELMANTEL, M. (2020): *Landeskunde im Kontext: Die Umsetzung von theoretischen Landeskundeansätzen in DaF-Lehrwerken*, Göttingen: V+R unipress.
- FALK, S. & GÖTZ, S. (2016): „A case study based on Duolingo“, in: ZEYER, T., STUHLMANN, S. & JONES, R. (Hrsg.): *Interaktivität beim Fremdspra-*

- chenlehren und -lernen mit digitalen Medien: Hit oder Hype?*, Tübingen: Narr: 237–259.
- FUCHS, O. (2012): *Entwicklung interkultureller Kompetenz in Deutsch als Fremdsprachenunterricht*. Dissertation, abrufbar unter: <https://dnb.info/1042303509/34>, 27.07.20.
- GREIN, M. (2018): „Progress in language teaching. From competence to dialogic competence-in-performance“, in: WEIGAND, E. & KECSKES, I. (Hrsg.): *From Pragmatics to Dialogue Cross-cultural pragmatics in language learning and teaching*, Amsterdam: John Benjamins: 61–82.
- GREIN, M. (2019): „Interkulturelle Kompetenz für DaF-Lehrende – Dialogisches Handlungsspiel und Multikollektivität“, in: ERSCH, C. M. (Hrsg.): *Kompetenzen in DaF/DaZ*, Berlin: Frank & Timme: 125–145, abrufbar unter: <https://library.oapen.org/handle/20.500.12657/24334>, 21.10.2020.
- GODWIN-JONES, R. (2017): „Smartphones and language learning“, in: *Language Learning & Technology* 21/2: 3–17, abrufbar unter: <http://llt.msu.edu/issues/june2017/emerging.pdf>, 27.07.20.
- HECKE, C. (2010): *Visuelle Kompetenz im Fremdsprachenunterricht: Die Bildwissenschaft als Schlüssel für einen kompetenzorientierten Bildeinsatz*. Dissertation, abrufbar unter: <https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-000D-EF96-D>, 30.07.20.
- LAI, C. & ZHENG, D. (2018): „Self-directed use of mobile devices for language learning beyond the classroom“, in: *ReCALL* 30/3: 299–318, abrufbar unter: <https://www.cambridge.org/core/journals/recall/article/selfdirected-use-of-mobile-devices-for-language-learning-beyond-the-classroom/C60A8CE7FA2F2D3CC4548398BDAB828B>, 30.07.20.
- ROSELL-AGUILAR, F. (2017): „State of the App: A Taxonomy and Framework for Evaluating Language Learning Mobile Applications“, in: *CALICO Journal* 34/2: 243–258.
- RÖSLER D. (2012): *Deutsch als Fremdsprache. Eine Einführung*, Stuttgart: J.B. Metzler.
- USENER, J. (2016): *Lehrwerke und interkulturelle Kompetenz im Spanischunterricht. Analyse und Perspektiven*. Dissertation, abrufbar unter: <http://digital.bibliothek.uni-halle.de/ulbhalhs/urn/urn:nbn:de:gbv:3:4-16923>, 30.07.20.

- USTA, M. (2017): *Fremdsprachenlernen mit Apps? Eine Evaluation der App Duolingo*, abrufbar unter: <https://d-nb.info/116133940X/34>, 27.07.20.
- WARD, M. (2020): *Language app Duolingo has seen a 148% spike in sign-ups. Here are the 5 most popular languages people in self-isolation are learning right now*, abrufbar unter: <https://www.businessinsider.com/duolingo-sees-spike-in-new-users-during-coronavirus-pandemic-2020-4?r=DE&IR=T#5-german-for-english-speakers-1>, 19.04.2020.
- WITZENLEITER, H. & LUPPOLD, S. (2020): *Quick Guide Interkulturelle Kompetenz: Interkulturelle Sensibilisierung für eine grenzenlos erfolgreiche Kommunikation Quick Guide*. Wiesbaden: Springer.

## Internetquellen

- BABEL: Lernen mit Babbel, abrufbar unter: <https://about.babbel.com/de/>, 27.07.20.
- BUSUU: So funktioniert Busuu, abrufbar unter: <https://www.busuu.com/de/it-works/courses>, 30.07.20.
- DEUTSCHE GESELLSCHAFT FÜR VERBRAUCHERSTUDIEN: Sprachlernportale: Angebot & Preise, Lernspaß, Lernkomfort, Transparenz, Umfang der App, abrufbar unter: <https://www.dtg.de/tests/sprachlernportale-2018/>, 30.07.20.
- DUOLINGO: Duolingo, abrufbar unter: <https://de.duolingo.com/>, 30.07.20.
- GOOGLE PLAY STORE: Apps. Sprache lernen, abrufbar unter: <https://play.google.com/store/search?q=sprachen%20lernen&c=apps>, 27.07.20.
- YOUTUBE: Busuu. 12 German greetings they WON'T teach you in school (Native Speaker), abrufbar unter: <https://www.youtube.com/watch?v=flfayo4Hij0&list=PL1Dxz9WLgT4h6PV7W2-oMSAgm99egJ8rB&index=3&t=0s>, 27.07.20.



## **Geeignet zum DaF/DaZ-Lehren? Evaluation der Rolle sozialer und personaler Kompetenzen in Eignungstests für DaF/DaZ-Studierende**

### **Einleitung**

Als ich mich letztes Jahr bei einem DAX-Unternehmen<sup>1</sup> als Trainerin bewarb, absolvierte ich dafür ein dreistufiges Assessment Center<sup>2</sup> bestehend aus einem einstündigen psychologischen Eignungsgespräch (inklusive Rollenspielen), einer fremdbeobachteten Arbeitssimulation, die die Vorbereitung und Durchführung einer Unterrichtseinheit umfasste, und einem Gruppengespräch. Laut Ausschreibung ging es bei diesem Auswahlverfahren vor allem darum, meine sozio-kommunikativen Kompetenzen und Fähigkeiten als Motivator\*in zu evaluieren/testen, denn diese wurden als Schlüsselkompetenzen für die Trainings-Aufgabe eingestuft.

In Organisationen der freien Wirtschaft sind eignungsdiagnostische Verfahren dieser Art übliche Praxis. Der Personalauswahl und -entwicklung wird große Bedeutung für den langfristigen Erfolg des Unternehmens beigemessen und entsprechend werden viele Ressourcen und Aufwand in die Prüfung der berufsrelevanten Persönlichkeitsmerkmale und Kompetenzen investiert (vgl. Hechinger & Seibert 2015: 42). Sozialen und personalen Kompetenzen wie

.....

- 1 Aus Gründen der vertraglich vereinbarten Geheimhaltung wird der Name des Unternehmens anonymisiert.
- 2 Ein Assessment Center (AC) ist eine Methode der Personalauswahl und -entwicklung, die eine Sequenz von Situationen darstellt, in denen die Teilnehmer\*innen alleine oder in Kleingruppen zusammen mit anderen Teilnehmer\*innen und Beobachter\*innen Aufgaben bearbeiten, die Aufschluss über arbeitsbezogene Fähigkeiten, insbesondere soziale Kompetenzen, geben sollen (Fisseni & Fennekels 1995; Hossiep 1994 in Asendorp & Neyer 2012: 164). Hierzu zählen auch durch Expert\*innen beobachtete tätigkeitsnahe Simulationen und Rollenspiele, aber auch Tests, persönliche Interviews oder Gruppengespräche (Panels).



Konfliktfähigkeit, Durchsetzungsvermögen, kommunikative Kompetenz und Empathie werden dabei eine immer größere Wichtigkeit zugeschrieben und mithilfe berufspsychologischer Methoden wie o.g. ACs, insbesondere bei Führungsaufgaben, intensiv getestet (vgl. Schmidt 2016). Denn diese Fähigkeiten werden als „Humankapital“ betrachtet (vgl. Brohm 2009: 62).

In personenzentrierten Berufen (wie dem Lehrberuf) gelten soziale Kompetenzen sogar als Schlüsselkompetenzen (vgl. Kanning 2005: 14; Boeger 2016: 6). Erstaunlicherweise finden diese jedoch selten eine explizite Beachtung in Kompetenzmodellen für Lehrende wie z.B. das *europäische Profilraster für Sprachlehrende* (vgl. EPR 2020).

Eine vergleichbare Prüfung der persönlichen Eignung angehender Lehrenden vor der Studienaufnahme wird erst in den letzten Jahren kontrovers diskutiert und ansatzweise getestet (vgl. Boeger 2016; Mayr 2015; Kieschke 2015; Nieskens 2016; Schaarschmidt et al. 2017, Klusmann et al. 2011; Renger et al. 2019). Als Auslöser der Debatte um Eignungsprüfungen gilt u.a. die Unterschätzung der Lehrtätigkeit und ihrer Herausforderungen (vgl. KMK 2013: 2; siehe dazu auch Kieschke 2015: 23). Auch die Erfahrungen aus Finnland, wo selektive Eignungsprüfungen in Form von ACs durchgeführt werden, spielen nicht zuletzt wegen des bildungspolitischen Erfolgs Finnlands eine Referenzrolle (vgl. Kricke 2016: 305; Mayr 2015: 62). In Deutschland hat der Einsatz eignungsdiagnostischer Verfahren wie des *Parcours*<sup>3</sup> oder des *CCT*<sup>4</sup> bisher

.....  
3 *Parcours*: Eintägiges AC der Universität Passau mit Simulation lehramtsspezifischer Situationen zur Fremdeinschätzung der Lehreignung. Das Verfahren umfasst Selbstpräsentation, Gruppendiskussion, Empathieübungen und Rollenspiele. Das Verfahren ist primär als Beratungsangebot und nicht zur Fremdselektion konzipiert (Seibert & Hechinger 2009 in Renger 2019: 39).

4 *Career Counselling for Teachers* (CCT) ist ein onlinebasiertes Selbsterkundungsangebot zur Selbsteinschätzung der persönlichen Eignung und Motivation für den Lehrberuf und das grundsätzlich als Beratungsprogramm von Lehrbildungsausrichtungen aus sechs europäischen Ländern konzipiert wurde. Die Online-Fragebögen dienen der Auseinandersetzung mit persönlichen Einstellungen, Fähigkeiten und Erfahrungen und ihre Bedeutung für die Profession mit Ergebnismeldungen. Ferner umfasst CCT auch geführte Touren für die Interpretation der Ergebnisse, Informationen über das Bildungswesen und den Lehrberuf sowie Karrieremöglichkeiten und Reportagen von Lehramtstudierenden und Lehrenden mit Berichten aus ihrem Leben (vgl. Mayr 2015: 60; <https://www.cct-germany.de/CCT/SetAudience>).

keine Konsequenz für die Zulassung zum Studium (vgl. Boeger 2016: 2). Der Trend in der bildungspolitischen Diskussion verlangt jedoch zunehmend nach selektiven Auswahlverfahren unter Berücksichtigung berufsrelevanter Persönlichkeitsmerkmale und Interessen sowie psychosozialer Kompetenzen (vgl. Nieskens 2016: 155; Schaarschmidt et al. 2017: 43–47; Mayr 2015: 56, 58; Renger et al. 2019: 52).

Während die Relevanz sozialer Kompetenzen in der Berufswelt allgemein anerkannt ist und für das Lehramt zumindest diskutiert wird, findet im DaF/DaZ-Bereich eine vergleichbare Eignungsdiskussion bisher nicht statt. Dies erscheint erstaunlich, denn zu den Herausforderungen des regulären Unterrichts stellen Faktoren wie die enorme Heterogenität und soziale Komplexität in den DaF/DaZ-Klassenzimmern besondere psychosoziale Anforderungen an die Lehrkräfte, deren Bewältigung ausgeprägte soziale und personale Kompetenzen erfordert (vgl. Tissot et al. 2019: 73f.).

Diese erstaunliche Diskrepanz führte zu den Fragen: Wie weit ist der DaF/DaZ-Bereich in Bezug auf Eignungsdiagnostik im Vergleich zum Lehramt und zur freien Wirtschaft? Findet eine Prüfung der sozialen und personalen Kompetenzen von Bewerber\*innen in den Auswahlverfahren der DaF/DaZ-Bildungseinrichtungen bereits Eingang? Die hierzu durchgeführte Untersuchung wird in diesem Beitrag zusammenfassend dargestellt. Dabei werden zunächst die sozialen und personalen Kompetenzen umrissen (Kapitel 1). Im Anschluss daran wird ihre Relevanz für den Lehrberuf (Kapitel 2) und für den DaF/DaZ-Bereich evidenzbasiert dargestellt sowie ein Modell sozialer und personaler Kompetenzen vorgeschlagen, das sich an den konkreten Anforderungen der DaF/DaZ-Lehrtätigkeit orientiert (Kapitel 3). Es folgt eine kurze Beschreibung des empirischen Designs (Kapitel 3) und eine Zusammenfassung der Untersuchungsergebnisse (Kapitel 4).

## **1 Soziale und personale Kompetenzen und ihre Erfassung**

Was unter sozialen und personalen Kompetenzen verstanden wird, ist nicht einheitlich. Dies liegt u.a. an der Fülle der Fähigkeitskonzepte, die den beiden

multidimensionalen Konstrukten zugeordnet werden und an der Vielzahl der synonymisch verwendeten Bezeichnungen wie z.B. soziale und emotionale Intelligenz (vgl. Goleman 2006).

Es besteht dennoch Konsens darüber, dass es nicht ein allgemeingültiges sozial kompetentes Verhalten gibt, sondern dass dieses stets situationsgebunden ist (vgl. Greif 1997 in Rothland 2010: 583; siehe dazu auch Kanning 2005: 5 und Brohm 2009: 89). Sozial kompetentes Verhalten ist deshalb nicht inhaltlich, sondern nur kontextabhängig zu definieren und an seiner Situationsangemessenheit zu bewerten (vgl. Hinsch & Pfungsten 2002: 82 in Brohm 2009: 63, Kanning 2005: 5).

In der Psychologie umfassen die sog. sozio-emotionalen Kompetenzen ein breites Spektrum an Schlüsselfähigkeiten, die Individuen helfen, eigene und fremde Emotionen zu erkennen und zu regulieren, für andere zu sorgen, verantwortungsvolle Entscheidungen zu treffen und effizient mit alltäglichen Herausforderungen umzugehen (vgl. Gkonou & Mercer 2018: 158). In Handlungskompetenzansätzen bezeichnen diese Kompetenzen die

Dispositionen, kommunikativ und kooperativ selbstorganisiert zu handeln, d.h. sich mit anderen kreativ auseinander und zusammensetzen, sich gruppen- und beziehungsorientiert zu verhalten und neue Pläne, Aufgaben und Ziele zu entwickeln (Erpenbeck 2017: XXV).

Während soziale Kompetenzen eher Fähigkeitskonzepte umfassen, die in der sozialen Interaktion von Belang sind (z.B. Einfühlungsvermögen, Kontaktbereitschaft oder Durchsetzungsfähigkeit), fokussieren personale Kompetenzen eher die intrapersonale, subjektive Dimension sozialer Kompetenz wie Selbststeuerung, Motivation oder Reflexionsfähigkeit (vgl. Brohm 2009: 61). Personale Kompetenzen bezeichnen demnach die Fähigkeit und Bereitschaft einer Person,

reflexiv selbstorganisiert zu handeln, d. h. sich selbst einzuschätzen, produktive Einstellungen, Werthaltungen, Motive und Selbstbilder zu entwickeln, eigene Begabungen, Motivationen, Leistungsvorsätze zu

entfalten und sich im Rahmen der Arbeit und außerhalb kreativ zu entwickeln und zu lernen (Erpenbeck 2017: XXV).

Analog zu anderen Kompetenzen umfassen soziale und personale Kompetenzen sowohl kognitive Dispositionen, Wissen und Fähigkeiten, die zum kompetenten Verhalten befähigen, als auch affektive Elemente wie Werte, Emotionen, Motivation, Einstellungen und Bereitschaften, die als Performanzfaktoren dieser Fähigkeiten fungieren (vgl. Brohm 2009: 89).

Wie dies bei anderen Kompetenzen auch der Fall ist, stellt dieses Verhältnis zwischen den kognitiven, den motivationalen und den volitionalen Konstituenten von sozialer Kompetenz und ihren jeweiligen potenziellen kontextbezogenen Variationen eine Herausforderung für ihre Erfassung dar (vgl. Zlatkin-Troitschanskaia & Seidel 2011: 228–229). Diese Kontextabhängigkeit zusammen mit der Multidimensionalität der Konstrukte und die Schwierigkeit zu bestimmen, welche Kompetenzbereiche stabile oder welche veränderbare Persönlichkeitsmerkmale umfassen, macht ihre Erfassung denkbar schwieriger als die Messung von kognitiven Leistungen (vgl. Erpenbeck 2017: XIX; vgl. Zlatkin-Troitschanskaia & Seidel 2011: 226). Soziale und personale Kompetenzbereiche werden deshalb auch als *soft skills* oder weiche Faktoren bezeichnet. Diese Bezeichnung deutet in Abgrenzung zu *hard skills*, die die fachlich-technischen, messbaren Fähigkeiten für einen bestimmten Beruf umfassen, auf die schwierigere Erfassbarkeit und Erlernbarkeit der damit gemeinten übergreifenden Fähigkeitsbereiche hin (vgl. Erpenbeck 2017: XIX).

Eine Selektion der Studierenden nach weichen Kriterien wird deshalb mit Skepsis betrachtet. Kritisiert wird häufig die eingeschränkte diagnostische und prognostische Validität der eignungsdiagnostischen Instrumente (vgl. Klusmann et al. 2011: 718; Hechinger & Seibert 2015: 42, Asendorp & Neyer 2012: 163). Insbesondere Auswahlverfahren, die auf Selbsteinschätzungen (sog. Self-Assessments) zurückgreifen und in erster Linie Persönlichkeitsmerkmale berücksichtigen, werden wegen der geringen Validität kritisiert (vgl. Blömecke 2009: 105 in Bosse 2012: 86). Die für den Lehrberuf relevanten sozialen und personalen Kompetenzen ließen sich nur im persönlichen Kontakt (sog. *face-to-face-Assessments*) oder durch Verhaltensbeobachtung (z.B. durch eine Fremdeinschätzung im AC) valide erfassen (vgl. Renger et al. 2019: 49).

Die Schwierigkeiten ihrer Erfassung mindern jedoch nicht die Relevanz dieser Kompetenzen für den Lehrberuf. Dafür sprechen zahlreiche empirische Befunde, die im Folgenden erläutert werden.

## **2 Die Rolle der Lehrkraft und ihrer sozialen und personalen Kompetenzen**

Spätestens seit der Hattie-Studie besteht empirische Evidenz über die große Bedeutung der Lehrperson und vom Einfluss positiver Lernende-Lehrende-Beziehungen in Lernprozessen (vgl. Hattie 2009: 238, 297). Seitdem die Hirnforschung die Verbindung von Emotion und Kognition und damit die tragende Bedeutung von Beziehungen und affektiven Faktoren für die Verankerung von Wissen erklären und anhand der Aktivität in den Gehirnregionen messen kann, ist die wichtige Rolle der Lehrkraft in Lernprozessen auch wissenschaftlich nachweisbar (vgl. Hüther 2004: 31ff.).

Die sozialen Kompetenzen übernehmen bei der Lehrtätigkeit allein deshalb eine bedeutende Rolle, weil der Unterricht an sich für eine ständige soziale Interaktion steht. Als besonders erfolgreich gelten dabei solche Lehrende, die einerseits eine instruierende „Regisseurrolle“ im Unterricht übernehmen und als „Aktivatoren“ agieren (vgl. Hattie 2009: 243), und andererseits in der Lage sind, eine Perspektivenübernahme zu vollziehen (vgl. ebd. 2009: 238). Empathie und Führungsfähigkeiten erscheinen im Unterrichtskontext als scheinbar gegensätzliche, aber einleuchtend notwendige soziale Kompetenzen für die Bewältigung der Anforderungen. Nach der Metaanalyse von Wang, Härtel & Walberg wird die als „Gruppenmanagement“ bezeichnete Fähigkeit, Lehr- und Lernprozesse in einem Kollektiv erfolgreich zu organisieren, Beziehungen zu vielen Menschen gleichzeitig zu gestalten und ihnen das Gefühl zu geben, der Unterricht sei für sie persönlich gestaltet, als der effektstärkste Faktor von gutem Unterricht identifiziert (vgl. 1993 in Dollase 2013: 89).

Im Fremdsprachenbereich gelten soziale Kompetenzen als besonders günstige Faktoren für den Lernprozess wegen der Bedeutung moderner kommunikativer Lern- und Lehrmethoden im Spracherwerb und wegen der interkulturellen, sozialen und interpersonellen Natur von Kommunikation (vgl.

Gkonou & Mercer 2017: 8). Diese Fähigkeitsbereiche begünstigen nachweislich die Gruppendynamik, die Klassenführung und die Qualität der Lehrende-Lernende-Beziehung (vgl. Patti 2006 und Reissmann 2006 in Gkonou & Mercer 2018: 160).

Empirische Studien hierzu zeigen, dass Lehrende mit hohen Werten an emotionaler und sozialer Intelligenz fähiger sind, ein gutes Klassenklima zu schaffen, in dem es Lernenden besser gelingt, die Motivation zu behalten (vgl. Elias & Arnold 2006; Graziano et al. 2007; Nizielski et al. 2012 in Gkonou & Mercer 2017: 7), und in dem die Häufigkeit von Konflikten und Aggressivität reduziert wird (Gross & Levenson 1993, 1997; Richards & Gross 1999 in Gkonou & Mercer 2017: 7). Die Fähigkeit, Beziehungen erfolgreich zu gestalten, wirkt sich außerdem positiv auf das Wohlbefinden und die Gesundheit der Lehrenden aus. Es wurde festgestellt, dass das Scheitern vieler Lehrender aufgrund von Erschöpfungstendenzen, wie z.B. Burnout, seine Ursache eher in mangelnden Beziehungsgestaltungsfähigkeiten als in den kognitiven Anforderungen des Berufs findet (vgl. Kieschke 2015: 28). Personalen Kompetenzen, welche die eigene Wahrnehmung und Emotionsregulation begünstigen, werden eine entscheidende Rolle für die psychische Gesundheit und Zufriedenheit der Lehrenden zugeschrieben (vgl. Kieschke 2015: 29, 32; Schaarschmidt et al. 2017: 37f.; Abujatum 2007: 122 in Rothland 2010: 584). Denn diese helfen im Umgang mit der Erlebnisdichte, dem Entscheidungsdruck und möglichen im Unterrichtsalltag sich ergebenden Frustrationen (vgl. Schaarschmidt et al. 2017: 37–38). Solche Frustrationen können sich z.B. in der Auseinandersetzung mit dem Schicksal von Lernenden ergeben, von denen die Lehrbeurteilung häufig abhängt, oder im Umgang mit langwierigen Lernprozessen wie der Alphabetisierung (vgl. Tissot et al. 2019: 75).

Die Studien, die sich mit dem Einfluss der sozialen Kompetenzen auf den Zweitspracherwerb befassen, kommen außerdem zu dem Ergebnis, dass Lehrende mit hohen Werten an sozio-emotionaler Intelligenz über eine höhere Selbstwirksamkeit<sup>5</sup> verfügen (vgl. Moafian & Ghanizadeh 2009 in Gkonou &

.....  
5 Selbstwirksamkeit drückt das Vertrauen und die Überzeugung von Lehrpersonen aus, die Lernentwicklung ihrer Lernenden persönlich beeinflussen zu können. Lehrkräfte mit hoher Selbstwirksamkeitsüberzeugung setzen sich höhere Ziele, bürden sich mehr Arbeit auf, inves-

Mercer 2018: 160). Dieser wiederum wird ein hoher Einfluss auf den Enthusiasmus und das Engagement von Lehrenden, ihre Bindung an den Lehrberuf und sogar auf die Leistungen der Lernenden zugesprochen (vgl. Tschannen-Moran, Woolfolk Hoy & Hoy 1998; Ashton & Webb 1986; Ditton & Arnoldt 2004 in Lipowsky 2006: 55). Welche sozialen und personalen Kompetenzen speziell im Deutsch als Fremdsprache/Deutsch als Zweitsprache-Bereich von Bedeutung sind, wird im folgenden Kapitel erläutert.

### 3 Welche sozialen und personalen Kompetenzen brauchen DaF/DaZ-Lehrende? Ein Modell

Zu den Herausforderungen des allgemeinen Unterrichts kommen im DaF/DaZ-Klassenraum diverse Faktoren hinzu, welche die soziale Komplexität erhöhen und von denen hier insbesondere der Faktor der Heterogenität hervorgehoben wird. Diese Heterogenität äußert sich nicht nur in Hinblick auf die Herkunft, wie dies in Regelschulen und DaF-Kursen der Hochschulen häufig zu erleben ist, sondern auch auf andere Faktoren wie Bildungsstand und Alter (vgl. Tissot et al. 2019: 73) sowie Gesundheitszustand aufgrund möglicher psychischer Belastungen wie Traumatisierung infolge von Fluchtmigration (vgl. ebd.: 50). Weitere Herausforderungen z.B. in Integrationskursen bilden häufige Konflikte unter den Teilnehmenden sowie zwischen Lehrkräften und Teilnehmenden aufgrund sozio-kultureller, politischer und ideologischer Verschiedenheiten, die es von der Lehrkraft mit Sensibilität zu schlichten gilt (vgl. ebd.: 75). Von Lehrkräften wird dann erwartet, in konflikträchtigen Unterrichtssituationen einer Eskalation entgegenzuwirken und den Teilnehmenden ggf. Grenzen aufzuzeigen (vgl. ebd.: 66). Die Fähigkeit, sich in solchen Situationen durchzusetzen und dabei keine neuen Konflikte auszulösen, sondern im Gegenteil positive Beziehungen aufrecht zu erhalten, ist hier gefragt. Empathie

---

tieren mehr Zeit für Planung und Nacharbeitung des Unterrichts, probieren gerne Neues aus, glauben an das Voranbringen schwächerer Lernenden, nutzen das Feedback aus der Klasse für Verbesserungen des Unterrichts und zeigen eine höhere Identifikation mit dem Beruf (vgl. Lipowsky 2006: 55).

und Mitgefühl mit den Teilnehmenden oder die Fähigkeit, diese zum Lernen zu motivieren, aber auch andere affektive Faktoren, wie ein ausgeprägtes Interesse an der Tätigkeit und eine intrinsische Motivation, wurden bei der letzten Evaluation der Integrationskurse als zentrale *soft skills* von DaZ-Lehrkräften identifiziert (vgl. ebd.: 74).

In diesem Zusammenhang erscheint die besondere Rolle einer empathischen Fürsorglichkeit, die offenbar die Basis von prosozialem Verhalten und altruistischem Engagement bildet (vgl. Howe 2013, Krznaric 2014 in Mercer 2016: 96) ein wichtiges Werkzeug, um diesen Aufgaben gerecht zu werden. In der positiven Psychologie wird diese Form der Empathie wegen ihres Beitrags zur menschlichen Verständigung als Instrument mit großem Potenzial für radikale Veränderungen des Umfelds und sogar der Gesellschaft angesehen (vgl. ebd.). Auch Werte wie Respekt, Toleranz und Offenheit gegenüber anderen Kulturen, die für die transkulturelle Kommunikation entscheidend sind, werden durch diese Form der Empathie gefördert, wodurch auch die Kommunikation allgemein verbessert wird, die letzten Endes beim Erlernen von Fremdsprachen beabsichtigt wird (vgl. Gkonou & Mercer 2017: 8).

Wie andere Lehrkräfte auch, bilden DaF/DaZ-Lehrende eine soziale Schnittstelle zu unterschiedlichen Bezugsgruppen, die von Lernenden und Kollegen bis zu Vorgesetzten und Institutionen reichen. Ein Gelingen erfordert ein hohes Maß an kommunikativer Sensibilität und Rücksichtnahme einerseits, und andererseits die Fähigkeit, eigene Positionen, wenn erforderlich, offensiv durchsetzen zu können (vgl. Schaarschmidt et al. 2017: 34f.). Außerdem bedarf es an dazugehörigen impulsgebenden motivationalen Werten und Einstellungen, die die vorhandenen sozialen Fähigkeiten zum Vollzug führen (vgl. Brohm 2009: 89), z.B. eine soziale Orientierung, Interesse an Menschen und Kommunikation sowie eine ausgeprägte Offenheit und Prosozialität. Prosozialität bezeichnet hier das Interesse und die Bereitschaft, sich mit den Anliegen fremder Menschen zu befassen und den Wunsch, ihre Probleme mit dem Fokus auf integrative und kooperative Lösungen zu verstehen (vgl. Kanning et al. 2012: 146).

Aber nicht nur im DaZ-Klassenraum, sondern auch in der Auslandsgermanistik, für die Tätigkeit in bildungspolitischen Organisationen wie dem Goethe Institut oder den DAAD-Lektoraten bilden soziale und personale



Kompetenzen wie Belastbarkeit, Zuverlässigkeit, Verhandlungsgeschick, Organisationsvermögen und Teamfähigkeit zentrale Schlüsselkompetenzen, die in Verhandlungen und bei der Vermittlung zwischen Institutionen aus dem In- und Ausland nötig sind (vgl. DAAD 2020a). Ferner werden DaF-Lehrende mit ihrem Interaktionsverhalten und ihren sozialen Haltungen als „Repräsentanten“ deutscher Sprache und Kultur im In- und Ausland wahrgenommen. Was Lehrende über ihre sozialen Interaktionen und Haltungen vermitteln, z.B. Offenheit und Wärme, hat somit einen direkten Einfluss auf die positive (oder negative) Wahrnehmung des Faches (vgl. Schart 2014: 36, Hattie 2009: 108) und vielleicht, biographisch-perspektivisch gesehen, auch auf ein langfristiges Interesse an deutscher Sprache und Kultur. Da sich die sozialen Kompetenzen von Lehrkräften als wichtige Prädiktoren der Zufriedenheit von Lernenden erwiesen haben (vgl. Kanning & Gärtner 2008 in Rothland 2010: 583) und eine positive Wahrnehmung Deutschlands im Ausland im Sinne außenpolitischer Kultur- und Bildungsarbeit ist (vgl. DAAD 2020b), müssten künftige DaF/DaZ-Lehrende m.E. für die große Bedeutung und Wirkung ihrer sozialen Handlungen und Einstellungen im Unterricht sensibilisiert werden.

Im Umgang mit der enormen Diversität der DaF/DaZ-Welt ist zudem entscheidend, wie offen und flexibel Lehrkräfte die eigenen subjektiven Überzeugungen und „Normalitätsvorstellungen“ reflektieren und ob sie bereit sind, die Auseinandersetzung damit konstruktiv zu gestalten und in die eigene Entwicklung zu integrieren (vgl. Terhart 2015: 26). Offenheit und Reflexion tragen so maßgebend zur Professionalisierung bei (vgl. ebd.). Dabei bezeichnet Reflexionskompetenz die grundsätzlich offene Haltung und Bereitschaft, sich mit schwierigen Inhalten auseinanderzusetzen, diese aus verschiedenen Perspektiven zu betrachten und Annahmen und Grundsätze stets zu hinterfragen (vgl. Loughran 1996, Fishman 1998: 16, Calderhead 2006: 36, Hatton & Smith 1995: 3 in Roters 2012: 113f.). In der Fremdsprachendidaktik besteht Konsens darüber, dass es Lehrenden besser gelingt, guten Unterricht zu gestalten, wenn sie eine forschende, kritische Haltung gegenüber dem eigenen Tun einnehmen (vgl. Schart 2014: 48). Auch Hattie stellt fest, dass besonders erfolgreiche Lehrende vor allem diejenigen sind, „die gegenüber neuen Erfahrungen offenbleiben, aus Fehlern lernen und die das Feedback ihrer Lernenden als Lernmotor für sich selbst betrachten“ (vgl. 2009: 35).

Beim Vergleich von Kompetenzmodellen und Katalogen relevanter sozialer und personaler Kompetenzen für die Lehrtätigkeit (vgl. Frey & Jung 2011: 561, Kanning 2012: 143, Grein 2017: 229ff.) und unter Berücksichtigung wissenschaftlicher Erkenntnisse über den günstigen Einfluss bestimmter persönlichkeitsrelevanter Merkmale (vgl. Mayr 2015: 56) lassen sich immer wiederkehrende Fähigkeitskonzepte erkennen, die als Kern von sozialer Kompetenz bei Lehrenden identifiziert wurden und für deren Relevanz hinreichende empirische Befunde vorliegen. Da es die „eine“ soziale Kompetenz nicht gibt, sondern sie nur kontextabhängig zu definieren ist, wurden hier die wichtigsten Dimensionen sozialer und personaler Kompetenz für die spezifischen Anforderungen des DaF/DaZ-Bereiches in einem Modell zusammengetragen und in Anlehnung an Kannings Taxonomie nach der Qualität dieser Fähigkeiten eingeordnet (vgl. 2005: 30):

1. Alle perzeptiv-kognitiven Fähigkeiten, die die soziale Wahrnehmung schärfen und die richtige Analyse von emotionalen und sozialen Situationen ermöglichen. Diese sind hilfreich für die erfolgreiche Etablierung von Beziehungen und positive Beziehungen haben nachweislich eine förderliche Wirkung auf Lernprozesse (vgl. Hattie 2009: 297):  
*Sensibilität, soziales Bewusstsein, Wahrnehmungskomplexität und allen voran Empathie*
2. Alle affektiv-motivationalen Komponenten wie Interessen, Werte, Einstellungen und Motive, die mit einer ausgeprägt offenen und sozialen Orientierung zusammenhängen, da diese als Performanzfaktoren sozialer Kompetenz fungieren (vgl. Brohm 2009: 89) und die Gestaltung von Beziehungen und die Bindung zu Individuen motiviert und fördert:  
*Kooperationsbereitschaft, Interesse an Menschen und Kommunikation, soziale Orientierung, Flexibilität, Fürsorglichkeit, Kontaktbereitschaft, und allen voran Offenheit und Prosozialität*
3. Alle behavioralen Faktoren oder *skills*, die die Umsetzung eines guten Gruppenmanagements sowie die Organisation ermöglichen und bei Konflikten helfen:

*Führungskompetenz, Organisationsfähigkeit, Konfliktfähigkeit, Selbstregulation (z.B. Geduld) und allen voran verbale und nonverbale kommunikative Fähigkeiten und Durchsetzungsfähigkeit*

4. Alle perzeptiv-kognitiven Fähigkeiten und Kompetenzen, die einen kritischen und reflexiven Umgang mit den eigenen Handlungen und Einstellungen ermöglichen, da diese als Antriebsquelle der Lehrprofessionalität fungieren (vgl. Terhart 2015: 26) und als Schlüssel zum weiteren Ausbau notwendiger Kompetenzen gelten (vgl. Boeger 2016: 7):

*Kritikfähigkeit, Ambiguitätstoleranz und allen voran Reflexionskompetenz*

Werden all die oben genannten Kompetenzen und Einstellungen auf ein vereinfachendes „Grundgerüst“ reduziert, ergibt sich folgendes Modell für den DaF/DaZ-Bereich:

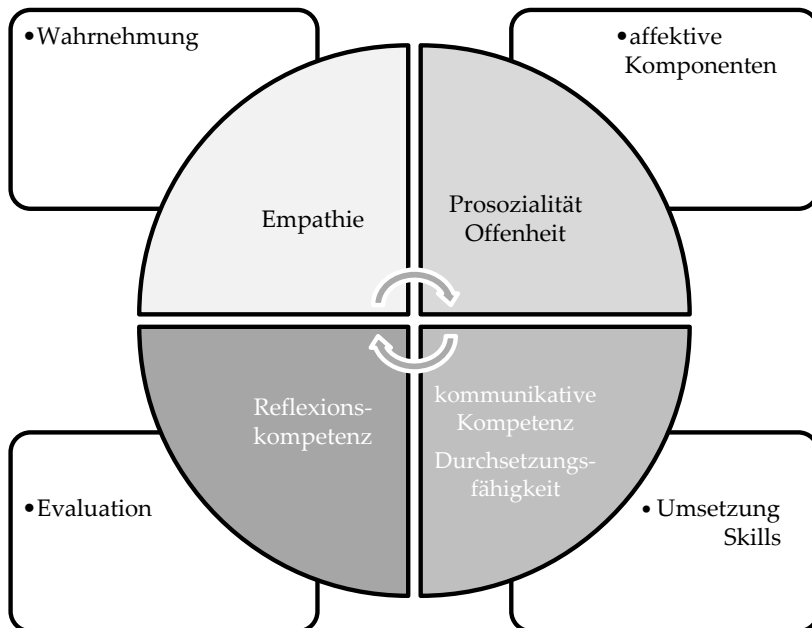


Abb. 1: Modell sozialer und personaler Kompetenzen für eine Lehrtätigkeit in DaF/DaZ-Bereich

Basierend auf der im Modell dargestellten sozialen und personalen Kompetenzen und im Kontext des zunehmenden bildungspolitischen Interesses an einer Eignungsprüfung künftiger Lehrender nach sozio-kommunikativen, personalen und motivationalen Kriterien (vgl. Nieskens 2016: 155), galt das Forschungsinteresse für vorliegende Untersuchung der Frage, inwiefern diese Voraussetzungen in den Auswahlverfahren der DaF/DaZ-Masterstudiengänge Berücksichtigung finden. Ferner sollte die Studie zeigen, wie DaF/DaZ-Bildungseinrichtungen gegenüber einem grundsätzlichen Einsatz von eignungsdiagnostischen Verfahren für die Prüfung von berufsrelevanten sozialen und personalen Kompetenzen vor der Studienaufnahme eingestellt sind. Im Folgenden wird der Aufbau der Studie skizziert.

## 4 Die Studie

Zur Datenerhebung wurden teilstrukturierte leitfadengestützte Expert\*innen-Interviews durchgeführt (vgl. Helfferich 2014: 561, 569). Für die Auswahl der Interviewpartner\*innen wurde die FaDaF-Zusammenstellung der in Deutschland angebotener DaF/DaZ-Studiengänge herangezogen (vgl. FaDaF 2018). Auswahlkriterium für die Befragung der Bildungseinrichtung war eine für den Zulassungsprozess zum Master DaF/DaZ vorgeschaltete Eignungsabklärung in Form einer persönlichen Begegnung, wie z.B. ein Eignungsgespräch oder ein Auswahlgespräch (*face-to-face-Assessments*). Da die Anzahl der diese Voraussetzungen erfüllenden Institutionen gering ausfiel<sup>6</sup> und nur vier der fünf relevanten Universitäten für ein Interview gewonnen werden konnten, wurden zum Vergleich auch Universitäten mit einer eingeschränkten Zahl an Studienplätzen kontaktiert, die Dokumente als Zulassungsvoraussetzung anforderten, die den Zugang zu Informationen über die Person und damit Rückschlüsse auf deren Einstellungen und Interessen ermöglichten, z.B. das Einreichen eines Motivationsschreibens und/oder eines Lebenslaufs. Insgesamt ergab sich eine Stichprobe von 5 Fällen. Davon führten vier Institutionen (B03–B06) Eig-

.....

6 Laut Internet-Recherche setzten ausschließlich fünf Institutionen ein Eignungsgespräch für die Zulassung voraus. Vollständigkeit dieser Angaben ohne Gewähr.

nungsgespräche im Rahmen ihrer Auswahlverfahren. Bei dem übrigen Fall konnten die Perspektiven zweier Expert\*innen erhoben werden (B01–B02).

Die für diese Untersuchung gewonnenen Expert\*innen waren allesamt Akteur\*innen von DaF/DaZ-Bildungseinrichtungen (fachliche Koordinator\*innen und Dozent\*innen bzw. Professor\*innen), die über fachliche Verantwortung für die Organisation der Studiengänge verfügten, an den Auswahlgesprächen beteiligt waren und/oder über die Kompetenz verfügten, über die Eignung der Bewerber\*innen zu entscheiden.

Die erhobenen Daten wurden nach der inhaltlich strukturierenden Inhaltsanalyse ausgewertet (vgl. Kuckartz 2018: 97ff.).

## 5 Ergebnisse

Das Forschungsprojekt wurde von der Annahme geleitet, dass potenziell affektiv geladene Unterlagen wie einem Motivationsschreiben und persönliche Interviews bzw. Auswahlgespräche, die im Rahmen von Auswahlverfahren zum Masterstudiengang DaF/DaZ stattfinden, zwingend zur Ermittlung der persönlichen Eignungsmerkmale konzipiert sein müssten, die den Anforderungen an die künftige DaF/DaZ-Lehrtätigkeit entsprechen. Damit sind vor allem die Aspekte der sozialen Interaktion und der persönlichen Orientierung sowie verbale und nonverbale kommunikative Fähigkeiten gemeint, die im Modell dargestellt wurden (s.o.). Die Gründe für eine persönliche Begegnung und den damit verbundenen hohen organisatorischen Aufwand seitens der Einrichtungen konnten, so die Annahme, ausschließlich in der Ermittlung jener Fähigkeiten und Kompetenzen liegen, die nicht ohne weiteres anhand herkömmlicher Bewerbungsunterlagen überprüft werden können und nur über ein *face-to-face*-Assessment oder eine Verhaltensbeobachtung einschätzbar und valide zu erfassen sind (vgl. Renger et al. 2019: 49). Die Ergebnisse, zeigten jedoch, dass andere Relevanzen für die Auswahl galten: Die geringe Gewichtung persönlicher Voraussetzungen zugunsten fachlicher Kriterien in den Auswahlverfahren und die hierfür angeführten Gründe werden jeweils in den Kapiteln 5.1 und 5.2 abgebildet. Ebenfalls wird die Rolle der Motivation für das Studium und praktische Erfahrungen thematisiert (Kapitel 5.3). In

Kapitel 5.4 wird näher auf die ambivalente Haltung der Institutionen gegenüber eignungsdiagnostischen Verfahren eingegangen, bevor auf die mutmaßlichen Konsequenzen für die Bewerber\*innenzahlen skizziert werden (Kapitel 5.5).

## 5.1 Persönliche Voraussetzungen: soziale und personale Kompetenzen

In Rahmen dieser Studie wurde festgestellt, dass die Entscheidung für oder gegen eine Zulassung den Angaben zufolge äußerst selten von persönlichen bzw. weichen Faktoren abhängt. Einige Befragte wiesen sogar ausdrücklich darauf hin, weiche Kriterien generell nicht vorauszusetzen (B01: 159–165; B02: 508):

*Vielleicht muss ich Sie nur gleich vorwarnen [...] es geht tatsächlich in diesem Interview, das wir führen, weniger darum zu gucken, ob die Leute... wie lehrgeeignet sie sind. (B04: 20–23)*

Als Rechtfertigung für die Außerachtlassung von Lehreignungskriterien im Auswahlverfahren gilt die Begründung, dass man andere Ziele als die Lehr-tätigkeit haben könne:

*Man könnte eben auch ein anderes Ziel haben, als Lehrer zu werden. Man könnte auch an der Uni bleiben nach dem DaF-Master. [...] Deshalb suchen wir danach nicht aus. (B02: 263–265)*

Erstaunlich erscheint dieses Ergebnis, wenn man bedenkt, dass soziale Kompetenzen wie Empathie, Offenheit, Durchsetzungsfähigkeit und eine berufsverbundene soziale Orientierung auf Nachfrage von der Mehrzahl der Interviewten als berufsrelevante Eignungsdimensionen betrachtet werden, um den spezifischen Anforderungen der DaF- oder DaZ-Lehr-tätigkeit, etwa durch die soziale Heterogenität der Lernenden, gerecht zu werden. Die Förderlichkeit personaler Kompetenzen wie Geduld (B04: 271), Belastungsfähigkeit (B03: 288–292), Selbstregulation und Resilienz (B04: 339–343) insbesondere bei DaZ-Kursen wurde dabei hervorgehoben:

*Ja, auch eine gewisse Resilienz. Dass man sich als Lehrperson – wie gesagt ... Gerade in Integrationskursen hat man ja oft mit schweren Schicksalen zu tun.* (B04: 339–340)

Die persönlichen Schilderungen aus der Lehrtätigkeit der Befragten B04 und B01/B02 lassen zudem schließen, dass die Anforderungen im DaZ-Bereich besondere persönliche Voraussetzungen, wie z.B. Selbstregulation und Durchsetzungsfähigkeit oder Mitgefühl erfordern (B04: 339–340, 329–335, 349–354), denn die Aufgaben ähneln hier häufig denen der Sozialarbeit (B02: 589–598), während diese bei Sprachvermittlungsaufgaben in lerngewohnten Gruppen nicht zwingend erforderlich sind. Die Aussagen hierzu vermitteln teilweise den Eindruck, der DaZ-Bereich leide deshalb bei einigen Befragten unter Status-Problemen und damit einhergehend unter einer eingeschränkten Bereitschaft, sich auf die besondere soziale Situation dieser Kurse einzustellen (B02: 458–461; B02: 589–598).

Im Bereich personaler Kompetenzen wird außerdem die besondere Rolle der Reflexionskompetenz betont, denn diese ist Voraussetzung für die kritische Betrachtung der eigenen persönlichen Eignung für den Lehrberuf (B03: 354–356) und für die Auseinandersetzung mit den eigenen Einstellungen und Überzeugungen (B05: 381–404). Reflexion wird zudem als wichtige Basis zur realistischen Selbsteinschätzung (B06: 464–467; B06: 472–473) und damit auch als Keim sozio-emotionaler Intelligenz und prosozialer Haltungen betrachtet:

*Für mich persönlich ist Reflexionskompetenz ganz wichtig. Ich muss mich erst mal selbst kennen, bevor ich andere kennenlernen kann und auch wertschätzen kann.* (B06: 472–473)

Bei der Aufzählung der Kompetenzen, über die man idealerweise im Hinblick auf eine Lehrtätigkeit in DaF/DaZ-Bereich verfügen sollte, wurden von den Interviewten mehrheitlich insbesondere folgende *soft skills* genannt: soziale Kompetenz (7 Nennungen), interkulturelle Kompetenz und Ambiguitätstoleranz (9 Nennungen), Empathie (6 Nennungen), Durchsetzungsvermögen (4 Nennungen), Prosozialität (4 Nennungen) und Offenheit (6 Nennungen):

*Tatsächlich eine gewisse Offenheit anderen Menschen gegenüber, also das was man vielleicht soziale Kompetenz nennen könnte, gehört ganz wesentlich dazu. Und das können wir im Studium tatsächlich nicht wirklich vermitteln, oder nur in sehr geringem Maße. (B05: 459–462)*

Demgegenüber steht das überraschende Ergebnis, dass eine gezielte Überprüfung dieser sozialen und personalen Kompetenzen jedoch nicht Gegenstand der meisten Auswahlgespräche ist, sondern nur in einem einzigen Fall nach vordefinierten Kriterien erfolgt.

5.1.1 Soziale Kompetenzen sind wichtig, werden aber kaum geprüft  
Einzig B06 gibt an, im Auswahlgespräch die persönlichen Voraussetzungen der Bewerber\*innen im Hinblick auf soziale Kompetenz und Eignung für den Lehrberuf formell zu prüfen (B06: 97–101; 183–184, 314–316, 326–328):

*[...] kann sie als Lehrkraft überzeugen? Hat sie eine offene Persönlichkeit? Also Offenheit ist uns sehr wichtig aufgrund unserer interkulturellen Situation. Ist sie flexibel? Oder kann sie... ist sie motivierend? [...]. Und dann wird das von 40 Punkten schon mit 10 Punkten bewertet. (B06: 190–194)*

*Wir schauen wirklich drauf: [...] kann diese Person den Beruf Lehrkraft erfüllen? Ist das Potenzial da? (B06: 383–385)*

Zu diesem Zweck steht B06 ein Kriterienkatalog zur Verfügung, der in die Blöcke „Interessen und Kompetenzen“ (B06: 92) sowie „Sprache, Argumentation und Persönlichkeit“ aufgeschlüsselt ist, für die es jeweils eine klar definierte Punktevergabe gibt (B06: 79–82, 131–136). Dabei werden im Eignungsgespräch die Bereiche Persönlichkeit, kommunikative Fähigkeiten und die Reflexion interkultureller Aspekte und praktischer Vorerfahrungen besonders fokussiert (B06: 108–118, 184–185, 326, 180–181, 198–199, 201–202, 326). Das Gespräch erfüllt vor allem die Funktion, ein Gesamtbild der Bewerber\*innen „als Person“ zu ermitteln (B06: 119–120; 374–377). Nach Angaben der Interviewten werden im Eignungsgespräch zahlreiche Dimensionen sozialer und personaler Kompetenz berücksichtigt und positiv bewertet (B06:



190–194). Hierzu zählen Empathie (B06: 202–203), Flexibilität (B06: 192,338), ein selbstbewusstes, motivierendes Auftreten (B06: 192; 316–318; 335–336), Durchsetzungsvermögen (B06: 326–328), Sensibilität für Interkulturelles (B06: 108–118, 201–202), Offenheit in Bezug auf Menschen und neue Erfahrungen (B06: 190–192, 409–412, 516–517) und Reflexionskompetenz (B06: 201–204, 328–329). Besonders die Reflexion möglicher Vorerfahrungen (B06: 161–166) und die Auseinandersetzung mit den eigenen Fähigkeiten und Defiziten im Hinblick auf eine Lehrtätigkeit werden abgefragt, um die Reflexionsfähigkeit der Kandidat\*innen zu testen (B06: 166–172). Belege für eine (pro)soziale Orientierung werden ebenfalls äußerst positiv bewertet (B06: 476, 511–518):

*Also was ganz wichtig bei uns auch ist: so was wie Ehrenamt. Wenn man sich im Ehrenamt engagiert hat, das zeigt für uns ja auch eine gewisse soziale Kompetenz. Das ist für uns eigentlich eher so das Wichtigste. (B06: 100–102)*

Zudem erfüllen die kommunikativen Kompetenzen der Bewerber\*innen eine gewichtige Rolle im Eignungsgespräch:

*Wie ist die Argumentation? Kann jemand Argumente formulieren? (B06: 178–181)*

*Glauben wir, dass die Lehrkraft überzeugend ist? (B06: 189–189)*

Der explizite Wunsch von B06 nach mehr Zeit für das Eignungsgespräch und besseren Instrumenten zur validen Erfassung der sozialen Kompetenzen macht deutlich, welche Wichtigkeit die Bildungseinrichtung diesen Faktoren beimisst:

*Natürlich wäre es toll, wenn man so einen ganzen Tag, einen Assessment Center Tag hätte... (B06: 354–355)*

Die kurze Zeit von 15–20 Minuten erlaube allerdings nicht, differenzierte Aussagen über die Bewerber\*innen zu treffen und verlässliche Entscheidungen über die Eignung der Bewerber\*innen für den Lehrberuf zu fällen (B06: 352–

354; B06: 332–333). Weitere Kritikpunkte beziehen sich auf die rudimentären verwendeten Mittel zur Eignungsdiagnostik (B06: 319–321).

Sinnbildlich für die Bedeutung sozialer und personaler Kompetenzen für diesen Studiengang steht der Erfahrungsbericht über einen Studierenden, der wegen einer betont arroganten Haltung, der Verwendung unangemessener Register und erheblichem Mangel an Interesse und sozialen Kompetenzen das Eignungsgespräch nicht bestanden hat (B06: 428–457). Dieser Erfahrungsbericht bestätigt jedoch auch, was sich m.E. aus dem ganzen Gespräch interpretieren lässt: Die Einrichtung würde das Eignungsgespräch im Zulassungsprozess gern höher gewichten, als dies aktuell der Fall ist:

*„Wir bedauern das ist die... die Entscheidung liegt nicht bei uns.“* (B06: 224–230)

Anders als bei B06 betonen die übrigen Interviewten, im Auswahlverfahren entweder soziale Kompetenzen überhaupt nicht zu prüfen (B01: 161–162, B02: 508, B04: 20–23) oder zumindest nicht dezidiert zu prüfen. Zwar gibt B05 an, ein „Mindestmaß“ an sozialer Kompetenz bei den Bewerber\*innen vorauszusetzen (B05: 435–437). Diese Aussage erfolgt allerdings erst nach ausforschenden Nachfragen der Interviewerin und nicht spontan bei der Frage nach den wichtigsten Faktoren in Auswahlgesprächen (B05: 286–294). Anders als bei B06 erfolgt die Prüfung des „Potenzials an sozialen, interkulturellen und Vermittlungskompetenzen, Offenheit und Flexibilität“ der Kandidat\*innen nicht gezielt und nach festen Kriterien, sondern besteht in einer eher situativen Bewertung. Nimmt man die zur Verfügung gestellten Erfahrungsberichte über Bewerber\*innen als Grundlage, entsteht der Eindruck, dass diese Kompetenzen und Einstellungen nicht positiv ermittelt werden, sondern erst ihr Fehlen negativ bewertet wird:

*Ich gehe in der Regel erst mal davon aus, dass Leute, die sich bei uns bewerben, eine gewisse grundsoziale Kompetenz mitbringen. Erst dann, wenn wir merken: da stimmt irgendwas nicht, dann wird man... kriegt auch schon ein bisschen große Ohren und kann dann auch nochmal gezielt dann nachhaken. Ja, das würden wir dann schon machen. Also spielt*

*eine Rolle, wenn jemand sich im Gespräch als sozial inkompetent herausstellt, weil er zum Beispiel überhaupt nicht auf Fragen eingeht. Dann wäre es ein Minuspunkt. Und ein ganz wichtiger. (B05: 494–500)*

Ähnliches geschieht im Fall B03. Auch hier lässt sich aus dem Gespräch nicht eindeutig erkennen, inwiefern persönliche Voraussetzungen als Auswahlkriterien berücksichtigt werden. Auf Nachfrage wird jedoch bei beiden Interviewpartner\*innen die enorme Bedeutung von Offenheit und Sensibilität für Interkulturelles insbesondere bei den binationalen Studiengängen bekräftigt (B03: 94–96; B05: 296–297; 299–302; 436–443; 302–309).

Im Fall B03 wird ebenso die Wichtigkeit kommunikativer Fähigkeiten angegeben, die über die überzeugende Darstellung einer Aufgabe (Vermittlungskompetenz) im Auswahlgespräch getestet werden (B03: 198; B03: 163–165). Personale Kompetenzen wie Organisationskompetenz (B03: 264) und Gewissenhaftigkeit (B03: 239–241) werden auf gezielte Nachfrage ebenfalls als wichtig angeführt. Ähnlich wie bei B05 bleibt jedoch nicht ermittelbar, welche Gewichtung diese Faktoren bei der Auswahlentscheidung haben und ob bzw. wie diese erhoben werden. Denn sie werden nicht aktiv als Auswahlkriterien formuliert, sondern erst nach Ausforschen der Interviewerin bezüglich der Kompetenzen, die in Hinblick auf die Lehrtätigkeit vom Belang sind, angesprochen.

### 5.1.2 Informelle Berücksichtigung der sozialen Kompetenzen mangels Erfassungsinstrumente?

In der kritischen Auseinandersetzung mit dem Interviewmaterial lassen sich allerdings Indizien identifizieren, die Rückschlüsse auf eine eventuell situative, informelle Berücksichtigung von weichen Kriterien zulassen:

*Eine Person, die sich auch Gedanken macht über das, was zum Beispiel Lernerinnen und Lerner gerade fühlen, gerade denken ... Was die Lernerinnen und Lerner betrifft, wenn sie diesen DaF/DaZ unterrichten [...] Und dass man einfach fähig ist, auch sich in andere Personen hineinzusetzen zu wissen: Das ist nicht das Übliche, was ich hier gerade sehe. Wenn sowas auch im Bewerbungsgespräch deutlich wird, ist das von Vorteil.*

*I: Würden Sie dann Empathiefähigkeit im Auswahlgespräch prüfen oder würden Sie einfach...?*

*B03: Wahrscheinlich nicht. Ich glaube nicht, dass man sowas prüfen kann. Da kann man fragen: „Bist du empathiefähig?“ Und dann sagt er: „Ja klaaaaar!“. So geht es nicht. Ich glaube, es ist auch nicht zielführend. Wir können keine Situation schaffen, in der wir das gründlich testen können. (B03: 296–309)*

Diese und ähnliche Aussagen z.B. über die Wichtigkeit von *soft skills* und von einer zugewandten Haltung der Kandidat\*innen (B04: 203–210; B05: 321–326; B05: 464–465) erlauben die Schlussfolgerung, dass die sozio-emotionalen und kommunikativen Kompetenzen der Bewerber\*innen möglicherweise nur aufgrund mangelnder Erfassungsmöglichkeiten nicht als Auswahlkriterien formuliert und nicht gezielt geprüft werden, jedoch bei positiven wie negativen Auffälligkeiten für die Entscheidung mitbewertet werden:

*I: Aber all diese [...] zum Beispiel soziale Kompetenz, würden Sie sie in einem Auswahlgespräch prüfen? [...]*

*B05: Jaaa, also sagen wir so [...] wenn wir, wenn also [...] ganz wichtig bei solchen Gesprächen egal welche Art von Gesprächen, auch Prüfungsgespräche, ganz wichtig ist: Habe ich den Eindruck, dass jemand mit dem ich spreche, überhaupt auf meine Fragen eingeht? (B05: 485–489)*

Es ist deshalb nicht auszuschließen, dass Beurteilungen nach dem Bauchgefühl erfolgen. Selbst die Bildungseinrichtung, deren Auswahlgespräche explizit auf die Prüfung der Lehreignung anhand personaler und sozialer Kompetenzen ausgerichtet waren, bedauerte, über keine ausgefeilten eignungsdiagnostischen Methoden für die Erfassung dieser Kompetenzen zu verfügen (B06: 319–321).

Ein weiteres interessantes Ergebnis erscheint mir in diesem Zusammenhang die ausdrückliche Betonung der Bewertungsobjektivität im Auswahlverfahren. Diese wird nach Angaben der Interviewten durch die Expertise erfahrener Personen oder durch die Beteiligung mehrerer Entscheidungsbefugter aus dem Auswahlgremium sichergestellt (B01: 85–91; B03: 172–174; B04: 105–106; B06: 185–189). In diesem Kontext wird zudem auf den ausschließlichen Bezug auf

klar definierte, harte Kriterien für die Auswahlentscheidung hingewiesen (B01: 393–395; B03: 172–174; B04: 125–128). Geht es um weiche Kriterien, wird die Verwendung eines transparenten Kriterienkatalogs betont, an dem sich die Entscheidung orientiert:

*I: Wie bilden Sie sich im Gespräch Ihre Meinung? Welche... Indikatoren bewerten Sie positiv?*

*B06: Also das hatte ich, glaube ich, auch erwähnt. Also wenn wir jetzt mal... Meinung im Sinne von... Was heißt Meinung? Wir haben ja unser Protokoll und daran orientieren wir uns. Das ist nicht nur eine Meinung, sondern ist unser Kriterienkatalog. (B06: 307–311)*

In Kontrast zu diesen Aussagen entsteht bei der Analyse der Aussagen ein zweites Bedeutungsmuster. Hier deuten zahlreiche sprachliche Indizien (glauben, denken, sehen, fühlen) auf eine Bewertung nach Bauchgefühl hin:

*[...] ich glaube schon, dass wir ganz gut nachfragen können und sehen können, ob eine Person sich auseinandergesetzt hat mit etwas oder eine Person nur auswendig gelernt hat, was sie gerade kann. (B03: 137–139)*

Inbesondere, wenn es sich um die Beurteilung von *soft skills* oder anderen affektiven Faktoren wie Motivation handelt, werden Intuition und Gefühl als Instrumente für die Entscheidungsfindung angegeben:

- *Also... ja, feste Indikatoren habe ich persönlich dafür nicht. Ich könnte Ihnen nicht auflisten, das und das und das. Es ist sehr individuell und es kommt sehr auf die individuelle und spezifische Gesprächssituation an. Man spürt das. (B05: 348–353)*
- *Haben wir das Gefühl, dass jemand Empathie hat und durch die Empathie diese Situation vielleicht bewerten kann? (B06: 202–203)*
- *Also die Person geht raus und die wollen wir haben. Weil wir einfach denken, das ist eine offene, interessierte, motivierte Person. Egal, ob jetzt sehr eloquent oder eher zurückhaltend... Einfach so. Es ist eher so ein...tatsächlich ein Gefühl. (B06: 411–413)*

- *Naja, wir wissen ja. Wir sind auch alle aus der Praxis und wir wissen ja auch, woraufes ankommt. Und ja, so versuchen wir, das zu bewerten [...] Wenn jemand motivierend erscheint und das kann man – glaube ich – auch nicht so richtig in Worte fassen – das ist halt vielleicht etwas Subjektives – das ist für uns auch schon auf jeden Fall ein Pluspunkt.* (B06: 334–338)

Die vermeintlich heikle Situation, aufgrund fehlender Kriterien für die Messung weicher Faktoren (B06: 352–354) nach dem Bauchgefühl entscheiden zu müssen, lässt es aus meiner Sicht plausibel erscheinen, dass einige Befragte die Benennung von weichen Faktoren durchgängig vermeiden und sich lieber hinter eine strikte Orientierung nach harten Kriterien zurückziehen. Dies zeigt sich aus meiner Sicht besonders prägnant beim wiederholten Kleinreden der weichen Kriterien im Auswahlverfahren von B01, wo selbst das Motivations-schreiben zur Überprüfung der sprachlichen Kompetenz dient:

*Das Motivationsschreiben ist vor allen Dingen auch ein Einblick zum einen, wie sich die Leute natürlich selbst präsentieren, und auch was für einen Text sie in so einer Situation verfassen können.* (B01: 139–141)  
*Absolut ungeeignet ist nach den harten Kriterien. Also wer das gar nicht erfüllt. Das ist sozusagen eine rein technische Sache.* (B01: 393–395)  
*I: Sie haben „Einstellungen“ genannt. Haben Sie was im Kopf, welche Einstellungen diese sein könnten?*  
*B01: Nicht auf Anhieb, weil das keine KO-Kriterien sind. Also es ist nicht das, was wir suchen.* (B01: 159–162)

### 5.1.3 Das Motivationsschreiben

Die Rolle des Motivationsschreibens (im Folgenden MS) wurde wegen seines potenziell affektiven Gehalts besonders in den Fokus dieser Untersuchung genommen. Anders als vermutet, dient das MS jedoch nicht, um Rückschlüsse auf die Person zu ermitteln, sondern vor allem dazu, die Gründe für die Studienwahl festzustellen (B05: 116–119) und damit auch das Interesse/den Enthusiasmus, mit dem die Studierenden ans Studium herangehen. Dies wird wegen

der Ressourcen, die Motivation für die Bewältigung des Studiums mobilisiert, als wichtig betrachtet: „*ob es so aussieht, dass sie das gut durchziehen können*“ (B01: 302–306), selbst wenn durch die Studiengänge ein erhöhter Workload zu erwarten ist, wie es bei Doppelstudierenden der Fall ist (B06: 40–47). Mögliche affektive Faktoren, die aus den Angaben im MS zu Rückschlüssen auf eine offene oder prosoziale Orientierung führen könnten, wie z.B. altruistische Gründe für die Studienwahl, werden nicht thematisiert. Damit konnte die Berücksichtigung sozialer Kompetenzen durch das Motivationsschreiben anhand der verfügbaren Daten nicht bestätigt werden.

## 5.2 Die eigentlichen Auswahlkriterien: Fachliche Voraussetzungen

Als Hauptziel aller Auswahlverfahren erweist sich vordergründig die Feststellung der Studierfähigkeit. Überraschend dabei ist, dass selbst die Interviews mit den Bewerber\*innen überwiegend auf die Überprüfung der fachlichen Kriterien und der auf Studienfach und -ort bezogenen Motivation konzipiert sind, um die Studierfähigkeit „realistisch einzuschätzen“.

In den Auswahlgesprächen werden mehrheitlich die Kenntnisse der Sprachwissenschaft und teilweise der Didaktik anhand von standardisierten Aufgaben zu Fehlerdiagnostik oder Syntexanalyse sowie durch Fragen zu Grammatik- und Sprachtheorien getestet (B03: 84–90; B04: 106–107; B05: 226–229). Die Sprachkompetenz wird in den Auswahlgesprächen ebenfalls gezielt und eingehend geprüft (B03: 324–327; B04: 128–129; B05: 71–72; B06: 54–56; 205–212).

In Zusammenhang mit diesem Ergebnis stehen nach Angaben der Interviewten vor allem wiederholt festgestellte Mängel an fachlichen und sprachlichen Kompetenzen unter den Studierenden. Ein rekurrentes Motiv sind die Berichte über negative Erfahrungen mit Bewerber\*innen aus dem Ausland. Insbesondere werden die für das Studium unzureichenden Deutschkenntnisse beklagt (B01: 144–149; B02: 400–401; B03: 166–169; B04: 231–237; B05: 71–74). Ebenfalls bemängelt werden die lückenhaften sprachwissenschaftlichen Grundkenntnisse der Bewerber\*innen aus dem In- und Ausland (B02: 278–279; B03: 324–331; B04: 144–146; B05: 230–234).

Die wiederholte Feststellung von Diskrepanzen und Abweichungen zwischen den durch Zeugnisse bescheinigten und den tatsächlich vorhandenen fachlichen und sprachlichen Kompetenzen der Studierenden erzeugt ein grundsätzliches Misstrauen der Institutionen gegenüber Sprachzertifikaten und sonstigen Zeugnissen (B03: 173–174; B04: 229–230; B05: 75–79; B05: 234–236), das sich deutlich auch an den Andeutungen über die Unechtheit bzw. dubiose Herkunft der Zertifikate erkennen lässt (B03: 166–168; B05: 75–79). Bedenkt man, dass so namhafte Sprachnachweise wie das *DSH3* oder der *TestDaF 4x4* für die Zulassung vorausgesetzt werden, erweist sich diese Problematik als besorgniserregend (B04: 72–76).

Die Bildungsinstitutionen reagieren auf diese Problematik mit einer eingehenden eigenhändigen Prüfung der Sprach- und Fachkompetenzen und richten zu diesem Zweck „Auswahlgespräche“ ein als eine Art mündlicher Prüfungsersatz. Die Durchführung von Auswahlgesprächen folgt demnach der Logik des Misstrauens und, nicht wie angenommen, der Ganzheitlichkeit. Drei von vier Institutionen setzen Auswahlgespräche deshalb vor allem ein, um die Zuverlässigkeit/Echtheit der eingereichten Zeugnisse zu überprüfen und die Angaben aus den Bewerbungsunterlagen besser „einschätzbar zu machen“ (B03: 136–140; B04: 72–78; B05: 51–53). Auch bei der Institution ohne Auswahlgespräch dienen die Angaben im Motivationsschreiben laut Aussagen der Befragten der Begutachtung der Sprachkompetenz (B01: 139–141). Einzig B06 verfolgt ein ganzheitliches Konzept, in dem versucht wird, ein gesamtes Bild der Bewerber\*innen und ihrer Kompetenzen durch das Gespräch zu erhalten (B06: 119–120).

### 5.3 Weitere Auswahlkriterien: Motivation und praktische Erfahrungen

Werden die sozialen Kompetenzen nur in einem Fall formell vorausgesetzt und geprüft, so erfüllen die persönlichen Voraussetzungen Interesse und Motivation eine ganz prominente Stellung als Auswahlkriterien bei fast allen Befragten (B03: 61–63; B03: 74–77; B04: 98–101; B05: 117–119; B06: 391–397; B06: 521–522). Vor allen Dingen interessieren sich die Institutionen für die Gründe der Studienwahl. Als motiviert gelten Bewerber\*innen, die nachweislich ein



ausgeprägtes Interesse an dem Fach DaF/DaZ zeigen, und zwar als Hauptfach (B01–B02: 359–366; B02: 337–340; B03: 198–200; B04: 101–105; B05: 145–150; B06: 150–153), und solche, die sich bereits in Voraus mit den spezifischen Inhalten und Dozenten der konkreten Bildungseinrichtung auseinandergesetzt haben (B05: 80–87; B05: 214–216; B06: 154–159). Bei binationalen Studiengängen wird dieses Interesse im Auswahlgespräch anhand des Bezuges zum Studienort (B03: 77–79) sowie einer überprüfbar eingehenden Befassung mit dem Zielland und seinen Besonderheiten gemessen (B03: 99–101; 143–147; 239–242; B05: 299–302).

Ausdrücklich betont wird auch die große Bedeutung von Praxiserfahrungen für die Auswahl. Denn diese erweisen sich den Angaben zufolge als Quelle wichtiger Erlebnisse, die mit dem Erwerb entscheidender Kompetenzen einhergehen (B03: 275–277; B03: 250–B05: 88–90; B05: 284–285; B05: 277–278; B06: 316–318). Einschlägige Erfahrungen in kultur- und sprachmittelnden Institutionen werden als festes Auswahlkriterium bei B05 mit 20% der Verfahrensnote am höchsten bewertet (B05: 120–121). Diese Erfahrungen werden wegen ihres Wertes als Wissensgrundlage geschätzt, da sie zu einem besseren Verständnis und zur Verinnerlichung der fachdidaktischen Inhalte im Master führen (B05: 155–166; B05: 170–173). Eine große Rolle bei der Auswahl spielen vor allem Unterrichtserfahrungen, denn diese zeugen zudem vom grundsätzlichen Interesse an der Lehrtätigkeit (B06: 160–162). Auslandserfahrungen gelten ebenfalls als lehrreich. Für die binationalen Master ist die Bedeutung von Auslandserfahrungen ebenfalls gewichtig für die Zulassung (B03: 275–277):

*Dass die Person schon einmal im Ausland war, finde ich auch sehr wichtig. Dieser Master DaF ist kein Auslandstudium, das man machen sollte, wenn man vorher nicht im Ausland war. Da sind wir uns einig. Weil hier vieles erwartet wird, was man einfach nicht weiß, wenn man nicht vorher im Ausland war. (B03: 254–257)*

Auch organisatorische Tätigkeiten in sozialen Einrichtungen oder Kulturinstitutionen werden als Ressource zum Ausbau relevanter personaler Kompetenzen, wie z.B. von Selbstorganisation (B03: 250–253) oder sozialer Kompetenzen, positiv bewertet (B06: 172–176).

Von den meisten Befragten wird zudem der Standpunkt geteilt, durch Erfahrungen ließe sich am ehesten feststellen, ob die Eignung zum DaF- bzw. DaZ-Lehren vorliegt (B01: 429–437; B02: 536–539; B03: 354–356; B06: 329–332; B06: 585–587). Außerdem könne man durch Erfahrungen automatisch zu den erforderlichen Kompetenzen gelangen (B02: 488–495; B02: 532–533; B03: 266–274; B03: 347–354; B04: 306–312; B06: 329–332), was gegen eine frühzeitige Selektion nach persönlichkeitsrelevanten Kriterien der Studierenden spreche:

*Das hat sich bei mir erst im Laufe meines ersten Studiums herausgestellt. Aufgrund eines Zufalls durch ein Auslandssemester, durch ein Praktikum am deutschen Institut habe ich gemerkt: HU! Deutsch zu unterrichten, macht ja total Spaß! Oder generell, dass ich gemerkt habe: Ich möchte gern mit Menschen arbeiten. (B06: 575–579)*

#### 5.4 Haltung gegenüber eignungsdiagnostischen Verfahren<sup>7</sup>

Bei den meisten Befragten lässt sich eine ambivalente Haltung gegenüber der Einführung eignungsdiagnostischer Verfahren vor der Studienaufnahme feststellen. Einerseits zeigen sich alle Befragten bis auf B01 mehr oder weniger dezidiert offen und positiv (B02: 237–239; B03: 206–208; B05: 513) bis hin zu enthusiastisch (B04: 384–385; B06: 354–355) gegenüber eignungsdiagnostischen Verfahren, die eine Feststellung der persönlichen Eignung für Studium und spätere Lehrtätigkeit durch die Prüfung der sozio-kommunikativen Kompetenzen und der berufsrelevanten Persönlichkeitsmerkmale zum Ziel haben. Die Idee einer Eignungsprüfung ausschließlich nach persönlichkeitsrelevanten Kriterien widerstrebt einzig B01:

.....  
7 Hierzu zählen Verfahren zur Eignungsfeststellung, die sowohl Eigeneinschätzungen (z.B. durch Fragebogen) als auch Fremdeinschätzungen/-beobachtungen (z.B. Auswahlgespräche, AC) beinhalten können.

*Ich weiß, dass es auf was anderes ankommt [...] da ist noch kein Wissen wirklich untergebracht. (B01: 556–560)*

Das Hauptargument für die Selektion von Bewerber\*innen mithilfe dieser Verfahren betrifft den großen Einfluss, aber auch die große Verantwortung, die Lehrkräfte im Hinblick auf die Lernenden haben:

*Das ist ja eigentlich eine super verantwortungsvolle Aufgabe, ja? Wissen zu vermitteln, Leute auf ihrem Weg zu begleiten, das ist Grundlage in der Schule, aber auch wenn wir neu Zugewanderte bekommen. Danach entscheidet sich deren ganzer Lebensweg so mit. Und man kann da unheimlich viel „reißen“. Man kann aber auch gleich viel kaputt machen. Und deswegen finde ich schon, dass es wichtig ist, dass man Lehrer gescheit auswertet oder auswählt und nicht jeden das machen lässt. (B04: 395–401)*

Weitere Argumente für eine Eignungsprüfung sind einerseits die willkommene Möglichkeit, dadurch an die Bestgeeigneten zu gelangen (B06: 359–362), und andererseits die beschränkte Aussagekraft, die eine Selektion ausschließlich nach Noten liefert:

*Und was zählt bei zulassungsbeschränkten Studiengängen ist: 1. ausschließlich die Abiturnote. Abiturnote sagt so wenig darüber, ob jemand ein guter Arzt wie darüber aus, ob jemand eine gute Lehrerin ist. Das ist ja völlig daneben. (B05: 513–517)*

Besonders positiv äußern sich zwei Befragte über einen möglichen Einsatz von Selbsterkundungs- und Selbsteinschätzungsverfahren, und zwar nicht als Selektionsinstrument, sondern als Mittel der Reflexion in unterschiedlichen Stufen der Lehrbildung (B02: 237–239; B02: 246–249; B06: 322–326; B06: 549–552).

Trotz der grundsätzlich positiven Einstellung der Akteur\*innen in Bezug auf die Eignungsdiagnostik lassen sich vier Kritik-Muster oder Kritik-Themenkomplexe erkennen, die mehr oder weniger durchgängig vertreten werden.

Größtes Problem aus Sicht der Befragten ist der Mangel (bzw. Unkenntnis/nicht Verfügbarkeit) an zuverlässigen Verfahren, die es erlauben würden, die ohnehin schwer greifbaren sozialen und personalen Kompetenzen der Bewerber\*innen valide zu erfassen (B01: 546–549; B03: 304–306; B05: 525–528, B06: 503–507). Hier wird exemplarisch die „Blenderproblematik“ bei Bewerber\*innen angeführt, die es verstehen, sich „gut zu verkaufen“ (B01: 546–549). Eine verlässliche diagnostische Validität der Verfahren wird deshalb auch als Bedingung für die Einführung solcher Verfahren angegeben.

*Von der Sache her finde ich es gut. Wenn man da ein aussagefähiges Kriterium oder Kriterien und auch Messverfahren entwickeln könnte und sie dann auch implementieren könnte und dann auch entsprechend Personal hätte, das zu machen, dann fände ich es wunderbar. (B05: 525–528)*

Es bestehen außerdem gewisse Vorbehalte, Verfahren aus der Berufspsychologie („Business-Bereich“) wie z.B. ACs für den Bildungsbereich einzusetzen (B02: 249–251; B05: 510–512). Allgemeine Skepsis herrscht auch in Bezug auf die eingeschränkten Möglichkeiten, verlässlich solche Fähigkeiten und subjektiven Einstellungen, wie z.B. Offenheit oder Empathie, zu messen (B03: 304–306; B05: 464–466; B06: 503–507).

Besonders kritisiert von B02 und B06 werden eignungsdiagnostische Verfahren, wenn sie als Ziel die Selektion verfolgen (B02: 239–241; B06: 552–553). Der frühe Zeitpunkt der Selektion in der Entwicklung könnte Personen aussortieren, die sich zu einem späteren Zeitpunkt als geeignet erweisen oder durch Ausbildung und Praxiserfahrungen zu „tollen Lehrkräften“ werden (B02: 244–246; B06: 555–564; B06: 575–579). Als großer Nachteil der Eignungsdiagnostik wird außerdem der hohe zeitliche und personelle Aufwand genannt (B03: 208–211; B05: 519–525; B06: 322–323).

## 5.5 Geringe Bewerber\*innen-Zahlen

Ein weiteres Motiv, das durch seine Rekurrenz in den Gesprächen auffällt, ist die Thematisierung der Bewerber\*innenzahlen bei allen Befragten. Das mag im Kontext von Auswahlverfahren selbstverständlich und trivial erscheinen,

jedoch lassen sich auch hier bedeutungstragende Aussagen ausmachen, die der Einführung eignungsdiagnostischer Verfahren entgegenstehen könnten. Vor allem die allgemeine Sorge um geringe bzw. sinkende Bewerber\*innenzahlen ist in vielen Gesprächen belegt (B01: 78–81; B02: 275–278; B04: 385–388; B06: 57; B06: 364–365). Für die meisten Befragten ist die aktuelle Zahl der Bewerbenden oder die Furcht vor schwindenden Bewerber\*innenzahlen sogar das einzige plausible Argument bzw. Gegenargument für oder gegen die Einführung von Eignungsprüfungen:

- *Ich glaube, die Frage kommt gar nicht auf, ob solche Verfahren bei uns angewendet werden. Ich glaube das... wir stehen da vielleicht gar nicht so dezidiert kritisch gegenüber, weil wir das noch nie mussten. Also die Frage hat sich nie gestellt bei uns.* (B01: 271–273)
- *Da kommt auch noch dazu, dass wir noch nie diese 30 Plätze überschritten haben. Wenn wir irgendwann überrannt würden, dann würden wir uns das vielleicht überlegen ((lacht)). Aber jetzt im Moment ist es eher noch so: Wir sind froh, wenn wir 30 gute Studierende haben* (B02: 275–278)
- *I: Könnten Sie sich vorstellen, ein solches Auswahlgespräch auch für den nationalen Studiengang, Masterstudiengang, einzuführen?*  
*B03: Könnte ich mir auch vorstellen. Da wir aber NC-frei sind, ist das unnötig* (B03: 190–192)
- *I: Sagen wir mal, wir hätten so viele Bewerber, wir hätten jedes Jahr 70 Bewerbungen. Würde sich dann anbieten, solche...?*  
*B04: Ja, spätestens dann, auf jeden Fall. Dann sehe ich keinen anderen Grund, der dagegenspricht.* (B04: 392–395)
- *Man muss aber einfach auch sagen, dass wir da ... Also es gibt sehr viel Konkurrenz auch zur Uni XY, was den Master DaF oder DaZ angeht, und sind uns dessen auch bewusst, dass jetzt... naja die Studierendenzahlen allgemein abnehmen und wir auch die Leute nicht zu sehr abschrecken wollen.* (B06: 355–359)

Die interpretative Generalisierung dieser Aussagen führt an dieser Stelle zu der These, dass die Akteur\*innen Eignungsprüfungen und Auswahlgesprä-

che eher als Barriere oder Filter auffassen und ähnlich wie das Festlegen von NC ausschließlich als Mittel zur Steuerung steigender Bewerber\*innenzahlen verstehen. Dabei werden die konstruktiven Möglichkeiten eignungsdiagnostischer Verfahren in den meisten Fällen verkannt, die als Gestaltungsinstrument kommender DaFlerInnen-Generationen genutzt werden könnten.

## 6 Schlussfolgerungen, Implikationen für die Praxis und Desiderate

Die subjektive, jedoch von praktisch allen Befragten geteilte Einschätzung, dass weiche Faktoren wie Empathie oder Offenheit zum DaF/DaZ-Lehren erforderlich sind, lässt an dieser Stelle den Schluss zu, dass alle Akteur\*innen die Bedeutung sozialer und personaler Kompetenzen als berufsrelevante Werkzeuge in DaF/DaZ-Lehrtätigkeiten anerkennen und somit als Faktoren der Lehreignung betrachten. Eine gezielte Prüfung der Lehreignung nach diesen Kriterien vor der Studienaufnahme erfolgt den Angaben zufolge jedoch in der Regel nicht bzw. nur in einem einzigen Fall. Bedenkt man, dass bei der Stichprobe fast alle DaF/DaZ-Bildungseinrichtungen vertreten waren, die ein *face-to-face*-Assessment für die Zulassung voraussetzen, kann behauptet werden, dass die sozialen und personalen Kompetenzen der Studienbewerber\*innen keine nennenswerte Rolle in den Auswahlverfahren zu DaF/DaZ-Masterstudiengängen spielen.

Vielmehr – und anders als angenommen – zielen die Auswahlgespräche mit den Bewerber\*innen überwiegend auf die Überprüfung fachlicher Kriterien wie Sprachkompetenz und Kenntnisse der Sprachwissenschaft, um die Studierfähigkeit realistisch einzuschätzen. Dies wird mit der Skepsis gegenüber den vorgelegten Sprachnachweisen und Zertifikaten begründet, die auf wiederholt negative Erfahrungen mit fachlich und sprachlich sehr schwachen Studierenden zurückgeführt wird. Eine eingehende Befassung mit dieser Problematik halte ich für notwendig, denn zahlreiche Aussagen der Befragten stellen ernsthafte Fragen an die Zuverlässigkeit zahlreicher anerkannter Sprachprüfungen wie *DSH*, *TestDaF* oder die Goethe-Zertifikate.

Obwohl bei Auswahlgesprächen mehrheitlich fachliche Kriterien geprüft wurden, gab es zahlreiche Anhaltspunkte dafür, dass auch Aspekte sozialer Kompetenz in die Bewertung einfließen, auch wenn diese nicht explizit als Auswahlkriterien formuliert wurden. Dies galt insbesondere, wenn Auffälligkeiten im Sozialverhalten erkannt wurden. Die Befragten wiesen in diesem Zusammenhang häufig auf das Fehlen geeigneter Instrumente zur validen Erfassung dieser Kompetenzen hin.

Um diese Lücke zu schließen, erscheint es m.E. sinnvoll, die Übertragbarkeit verschiedener, bereits bestehender eignungsdiagnostischer Verfahren für die Zwecke des DaF-/DaZ-Bereiches zu erforschen. Hierfür bieten sich die Instrumente zur Eignungsdiagnostik an, die für die Eignungsprüfung von Lehramtsstudierenden bereits vielerorts Anwendung finden. Hierzu zählen beispielsweise das *smk72* (vgl. Frey & Balzer 2017), *FIBEL* (vgl. Kanning et al. 2012), *FIT* (vgl. Schaarschmidt et al 2017), das bereits erwähnte CCT-Modul (vgl. Nieskens 2016) oder das Assessment Center Parcours (vgl. Hechinger & Seibert 2015). Diese Instrumente<sup>8</sup> bieten eine objektive Bewertungsgrundlage, auf die zur Überprüfung wichtiger Kriterien der Lehreignung zurückgegriffen werden könnte.

Eine solche Eignungsprüfung könnte aus meiner Sicht außerdem den positiven Effekt einer Aufwertung der DaF/DaZ-Studiengänge haben. Denn der hierfür betriebene Aufwand und die intensive Beschäftigung mit den persönlichen sozialen Fähigkeiten, Motivationen, Interessen und Einstellungen signalisieren den Bewerber\*innen zum einen, dass auch diese Kompetenzen für ihr Berufsziel erforderlich sind, und zum anderen, dass sie als Individuen wahrgenommen werden. Dass Auswahlgespräche tendenziell als wertschätzend empfunden werden, wurde auch durch einen Erfahrungsbericht aus dieser Untersuchung bestätigt (B06: 253–260). Den geäußerten Befürchtungen über einen abschreckenden Effekt eignungsdiagnostischer Verfahren (B04: 379–390; B06: 356–364) steht die Beobachtung zweier Befragter entgegen, dass Auswahlgespräche vor allem ungeeignete Bewerber\*innen abschrecken (B04: 129–135; B06: 45–461). Als Aufwertung könnten eignungsdiagnostische

.....

8 Eine Zusammenstellung der beim Lehramt eingesetzten eignungsdiagnostische Instrumente findet sich bei Renger et al. 2019.

Verfahren m.E. sogar umgekehrt einen positiven Einfluss auf die Bewerber\*innenzahlen haben.

Der eigentliche Nutzen solcher umfassenderen Eignungsprüfungen bei der Studienaufnahme aber bestünde aus meiner Sicht nicht nur darin, geeignete Studierende herauszufiltern, sondern vielmehr in der Möglichkeit, schon von Beginn an für die Wichtigkeit sozio-emotionaler Kompetenzen und sozialer/offener Einstellungen zu sensibilisieren. Außerdem könnten dabei bestehende Potenziale, Stärken und Defizite der Bewerber\*innen identifiziert werden, um sie zur kritischen Selbstreflexion nutzbar zu machen und während der Ausbildung systematisch zu bearbeiten. Beruflich-perspektivisch wären die so ausgebauten sozialen Kompetenzen unabhängig von der späteren Berufswahl ohnehin als Schlüsselqualifikationen vom Belang und auch im Hinblick auf ein lebenslanges Lernen für eine gesundheitsbewusste, zukunftsorientierte Lebensgestaltung wichtig (vgl. Amtsblatt der EU 2018: C219/4).

## Bibliografie

- AMTSBLATT DER EUROPÄISCHEN UNION (2018): *Empfehlung des Rates vom 22. Mai 2018 zu Schlüsselkompetenzen für Lebenslanges Lernen*, abrufbar unter: [https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32018H0604\(01\)&from=DE](https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32018H0604(01)&from=DE), 05.09.2020.
- ASENDORP, J. & NEYER, F. (2012): *Psychologie der Persönlichkeit*, Berlin: Springer.
- BOEGER, A. (2016): *Eignung für den Lehrerberuf. Auswahl und Förderung*, Wiesbaden: Springer.
- BOSSE, D. (2012): „Die Förderung ‚psychosozialer Basiskompetenzen‘ in der Lehrerbildung als Kontinuum gestalten“, in: DERS. et al. (Hrsg.): *Professionelle Lehrerbildung im Spannungsfeld von Eignung, Ausbildung und beruflicher Kompetenz*, Bad Heilbrunn: Julius Klinkhardt: 83–94.
- BROHM, M. (2009): *Sozialkompetenz und Schule. Theoretische Grundlagen und empirische Befunde zu Gelingensbedingungen sozialbezogener Interventionen*, Weinheim: Beltz.



- DOLLASE, R. (2013): „Lehrer-Schüler-Beziehungen und die Lehrerpersönlichkeit – wie stark ist ihr empirischer Einfluss auf Leistung und Sozialverhalten?“, in: KRAUTZ, J. & SCHIEREN, J. (2013): *Persönlichkeit und Beziehung als Grundlage der Pädagogik. Beiträge zur Pädagogik der Person*, Weinheim/Basel: Beltz: 85–94.
- ERPENBECK, J. et al. (2017): *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*, 3. Auflage, Stuttgart: Schäffer-Poeschel.
- FREY, A. & JUNG, C. (2011): „Kompetenzmodelle und Standards in Lehrerbildung und Lehrerberuf“, in: TERHART, E., BENNEWITZ, H. & ROTHLAND, M. (Hrsg.): *Handbuch der Forschung zum Lehrerberuf*, Münster: Waxmann: 540–572.
- FREY, A. & BALZER, L. (2017): „Einschätzungsbogen zu sozialen und methodischen Kompetenzen – smk72“, in: ERPENBECK, J. et al. (Hrsg.): *Handbuch Kompetenzmessung: Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*, 3. Auflage, Stuttgart: Schäffer-Poeschel: 149–160.
- GKONOU, C. & MERCER, S. (2017): *Understanding Emotional and Social Intelligence among English language Teachers*, London: British Council, abrufbar unter: [https://www.teachingenglish.org.uk/sites/teacheng/files/pub\\_G211\\_ELTRA\\_Gkonou%20and%20Mercer%20paper\\_FINAL\\_web.pdf](https://www.teachingenglish.org.uk/sites/teacheng/files/pub_G211_ELTRA_Gkonou%20and%20Mercer%20paper_FINAL_web.pdf), 25.06.2020.
- GKONOU, C. & MERCER, S. (2018): „The relational Beliefs and Practices of Highly Socio-emotionally Competent Language Teachers“, in: KOSTOULAS, A. & MERCER, S. (Hrsg.): *Language Teacher Psychology*, Blueridge Summit: Multilingual Matters: 158–177.
- GOLEMANN, D. (2006): *Soziale Intelligenz. Wer auf andere zugehen kann, hat mehr vom Leben*, München: Droemer.
- GREIN, M. (2017): „Das kompetenzorientierte Konzept der Deutsch-als-Fremdsprache-Ausbildung an der JGU Mainz“, in: BEHR, M. & SEUBERT, S. (Hrsg.): *Education is a Whole-Person Process – Von ganzheitlicher Lehre, Dolmetscherforschung und anderen Dingen*, Berlin: Frank & Timme: 223–238.

- HATTIE, J. (2009): *Visible Learning. A synthesis of over 800 meta-analyses relating to achievement*, London: Routledge.
- HECHINGER, M. & SEIBERT, N. (2015): „Eignungsverfahren für angehende (Gymnasial-)Lehrerinnen und Lehrer sind nötig?!“, in: KLITZING, S., DI FUCCIA, D. & STENGL-JÖRNS, R. (Hrsg.): *Auf die Lehrperson kommt es an? Beiträge zur Lehrerbildung nach John Hatties „Visible Learning“*, Kempten: Julius Klinkhardt: 38–66.
- HELFFERICH, C. (2014): „Leitfaden- und Experteninterviews“, in: BAUR, N. & BLASIUS, J. (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer VS: 559–574.
- HÜTHER, G. (2004): „Die Bedeutung emotionaler Reaktionen für Lernprozesse und die Verankerung neuer Erfahrungen“, in: BÖRNER, W. & VOGEL K. (Hrsg.): *Emotion und Kognition im Fremdsprachenunterricht*, Tübingen: Narr: 25–34.
- KANNING, U. (2005): *Soziale Kompetenzen. Entstehung, Diagnose und Förderung*, Göttingen: Hogrefe.
- KANNING, U., BÖTTCHER, W. & HERRMANN, C. (2012): „Measuring social competencies in the teaching profession – development of a self-assessment procedure“, in: *Journal for educational research online* 4/1, Münster: Waxmann: 140–154.
- KIESCHKE, U. (2015): „Zur Feststellung von Eignungsmerkmalen bei angehenden Lehrerinnen und Lehrern“, in: KLITZING, S., DI FUCCIA, D. & STENGL-JÖRNS, R. (Hrsg.): *Auf die Lehrperson kommt es an? Beiträge zur Lehrerbildung nach John Hatties „Visible Learning“*, Kempten: Julius Klinkhardt: 23–37.
- KLUSMANN, U., KÖLLER, M. & KUNTER, M. (2011): „Anmerkungen zur Validität eignungsdiagnostischer Verfahren bei angehenden Lehrkräften“, in: *Zeitschrift für Pädagogik* 57/5: 711–721.
- KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland] (2013): *Empfehlungen zur Eignungsabklärung in der ersten Phase der Lehrerausbildung. Beschluss der Kultusministerkonferenz vom 07.03.2013*, abrufbar unter: [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2013/2013-03-07-Empfehlung-Eignungsabklaerung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2013/2013-03-07-Empfehlung-Eignungsabklaerung.pdf), 05.09.2020.

- KRICKE M. (2016): „Die Auswahl von Lehramtstudierenden: Ein Beispiel aus Finnland“, in: BOEGER, A. (Hrsg.): *Eignung für den Lehrerberuf. Auswahl und Förderung*, Wiesbaden: Springer: 305–332.
- KUCKARTZ, U. (2018): *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, 4., überarbeitete Ausgabe, Weinheim u.a.: Beltz.
- LIPOWSKY, F. (2006): „Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler“, in: ALLEMANN-GHIONDA, C. & TERHART, E. (Hrsg.): *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern*, Weinheim u.a.: Beltz: 47–70.
- MAYR, J. (2015): „Zwischen Vielfalt und Kompetenz“, in: FISCHER et al. (Hrsg.): *Umgang mit Vielfalt*, Münster/New York: Waxmann: 55–75.
- MERCER, S. (2016): „Seeing the World Through Your Eyes: Empathy in Language Learning and Teaching“, in: MACINTYRE, P.D., GREGERSEN, T. & MERCER, S. (Hrsg.): *Positive Psychology in Second Language Acquisition*, Bristol: Multilingual Matters: 91–111.
- NIESKENS, B. (2016): „Eignungsabklärung und Zulassungssteuerung für den Lehrerberuf: Perspektiven, Instrumente und Erfahrungen“, in: BOEGER, A. (Hrsg.): *Eignung für den Lehrerberuf. Auswahl und Förderung*, Wiesbaden: Springer VS: 155–179.
- RENGER, S., KÖLLER, M. & KLUSMANN, U. (2019): „Eignungsdiagnostische“ *Verfahren für das Lehramtsstudium an deutschen Hochschulen – Überblick und Bewertung*, (Christian-Albrechts-Universität zu Kiel), abrufbar unter: [https://www.qualitaetsoffensive-lehrerbildung.uni-kiel.de/de/dateien-upload/broschuere\\_eignungsdiagnostische-verfahren-lehramt.pdf](https://www.qualitaetsoffensive-lehrerbildung.uni-kiel.de/de/dateien-upload/broschuere_eignungsdiagnostische-verfahren-lehramt.pdf), 05.09.2020.
- ROTTERS, B. (2012): *Professionalisierung durch Reflexion in der Lehrerbildung. Eine empirische Studie an einer deutschen und einer US-amerikanischen Universität*, Münster: Waxmann
- ROTHLAND, M. (2010): „Soziale Kompetenz: angehende Lehrkräfte, Ärzte und Juristen im Vergleich. Empirische Befunde zur Kompetenzausprägung und Kompetenzentwicklung im Rahmen des Studiums“, in: *Zeitschrift für Pädagogik* 56/4: 582–603.

- SCHAARSCHMIDT, U., KIESCHKE, U. & FISCHER, A. (2017): *Lehrereignung. Voraussetzungen erkennen, Kompetenzen fördern, Bedingungen gestalten*, Stuttgart: Kohlhammer.
- SCHART, M. (2014): „Die Lehrerrolle in der fremdsprachendidaktischen Forschung: Konzeptionen, Ergebnisse, Konsequenzen“, in: GNUTZMANN, C., KÖNIGS, G. F. & KÜSTER, L. (Hrsg.): *Der Fremdsprachenlehrer im Fokus*, Tübingen: Narr: 36–50.
- SCHMIDT, S. (2016): „So meistern Sie die häufigsten Übungen im Assessment Center“ in: *Süddeutsche Zeitung* vom 19.07.2016, abrufbar unter: <https://www.sueddeutsche.de/karriere/vorbereitung-fuers-auswahlverfahren-so-meistern-sie-die-haeufigsten-uebungen-im-assessment-center-1.3083273>, 05.09.2020.
- TERHART, E. (2015): „Heterogenität der Schüler – Professionalität der Lehrer: Ansprüche und Wirklichkeiten“, in: FISCHER et al. (Hrsg.): *Umgang mit Vielfalt*, Münster/New York: Waxmann: 13–30.
- TISSOT, A. et al. (2019): *Zwischenbericht zum Forschungsprojekt „Evaluation der Integrationskurse (EvIk)“*, (Bundesamt für Migration und Flüchtlinge), abrufbar unter: [https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Forschungsberichte/fb33-zwischenbericht-evik-I.pdf;jsessionid=AAB8A1933E28C281C5CA89D83E09C35E.internet572?\\_\\_blob=publicationFile&v=16](https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Forschungsberichte/fb33-zwischenbericht-evik-I.pdf;jsessionid=AAB8A1933E28C281C5CA89D83E09C35E.internet572?__blob=publicationFile&v=16), 05.09.2020.
- ZLATKIN-TROITSCHANSKAIA, O. & SEIDEL, J. (2011): „Kompetenz und ihre Erfassung – das neue ‚Theorie-Empirie-Problem‘ der empirischen Bildungsforschung?“ in: ZLATKIN-TROITSCHANSKAIA et al. (Hrsg.): *Stationen empirischer Bildungsforschung. Traditionslinien und Perspektiven*, Wiesbaden: Springer VS: 218–233.

## 7 Internetquellen

- DAAD (2020a): *Stellenausschreibungen für Lektorate und Sprachassistenzen*, abrufbar unter: <https://www2.daad.de/ausland/lehren/daadlektoren/de/16886-freie-lektorate-dozenten-und-sprachassistenzen/?s=1&projektid=57551600&page=1>, 05.09.2020.

DAAD (2020b): *Leitbild*, abrufbar unter: <https://www.daad.de/de/der-daad/wer-wir-sind/leitbild/>, 05.09.2020.

CAREER COUNSELLING FOR TEACHERS (2020): *Information und Selbsterkundung für den Lehrberuf*, abrufbar unter: <https://www.cct-germany.de/CCT/SetAudience>, 17.07.2020.

EPR (2020): *Europäisches Profilraster für Sprachlehrend*, abrufbar unter: <https://egrid.epg-project.eu/de/egrid>, 05.09.2020.

## Die medizinische Fachsprachenprüfung – Eine Pilotstudie zur Evaluation durch die Teilnehmenden

### Einleitung

[...] Die Gesundheit und das Wohlergehen meiner Patientin oder meines Patienten werden mein oberstes Anliegen sein [...]. Ich werde nicht zulassen, dass Erwägungen von Alter, Krankheit oder Behinderung, Glaube, ethnischer Herkunft, Geschlecht, Staatsangehörigkeit, politischer Zugehörigkeit, Rasse, sexueller Orientierung, sozialer Stellung oder jeglicher anderer Faktoren zwischen meine Pflichten und meine Patientin oder meinen Patienten treten [...]. (Auszug der Deklaration von Genf, Stand 2017)

Dieses Zitat ist ein Auszug aus der Genfer Deklaration, welche in Anlehnung an den hippokratischen Eid der 2. Generalversammlung des Weltärztebundes erstmals im Jahr 1948 verabschiedet wurde (vgl. Bundesärztekammer 2017). Es verdeutlicht die Haltung und Werte, nach welchen Mediziner\*innen handeln sollen und stellt eine unabdingbare Grundvoraussetzung für die Ausübung eines medizinischen Berufes dar. Eine gewissenhafte Ausübung und das Nachkommen der genannten ärztlichen Pflichten sind jedoch nicht ohne kompetente und fachspezifische Sprachkenntnisse sowie (inter)kulturelle Kompetenzen möglich. Dies ist im Besonderen für ausländische Ärzt\*innen oft eine Hürde (vgl. bspw. Schripf, Bahnemann & Lecher 2019: V).

In der Bundesrepublik Deutschland herrscht, neben einem allgemeinen Fachkräfteengpass im Gesundheits- und Pflegewesen (vgl. BfA 2019: 6), ein Ärzt\*innenmangel, wie auch die Zahlen der Ärztestatistik 2019 belegen (vgl. Bundesärztekammer 2019). Mediziner\*innen aus dem Ausland können diesen Bedarf decken (vgl. BAMF 2015: 5, 21f.) und Fachkräfte werden aus diesem

Grund gezielt angeworben (vgl. Hoesch 2003: 112) – dies bedeutet jedoch gleichzeitig, dass Maßnahmen für die sprachliche Ausbildung sowie zur Überprüfung der für die Berufsausübung notwendigen Sprachkenntnisse dieser Zielgruppe gegeben sein müssen.

Im Rahmen der 87. Gesundheitsministerkonferenz 2014 wurde bezüglich der Sprachkenntnisse ausländischer akademischer Heilberufe beschlossen,

[...], dass Personen, die in Deutschland in einem verkammerten akademischen Heilberuf tätig werden wollen, über ausreichende Kenntnisse sowohl der deutschen Umgangssprache als auch der medizinischen Fachsprache verfügen müssen [...] (2014: 2).

So sind fachspezifische Sprachkenntnisse nicht nur Voraussetzung für den (späteren) Berufsalltag als Ärztin oder Arzt, sondern der Nachweis dieser im Rahmen der Berufsankennung durch das Ablegen einer berufsspezifischen Fachsprachenprüfung (kurz: FSP) gesetzlich vorgeschrieben. Dies obliegt vornehmlich den Ärztekammern der Bundesländer. Aufgrund der kaum zugänglichen Datenlage in Bezug auf die Konzeption, den Ablauf etc. dieser Prüfungen seitens der Landesärztekammern, um festzustellen, ob es sich bei den vorhandenen Formaten um valide Nachweise medizinischer Sprachkenntnisse handelt, stellen die Prüfungsteilnehmenden (TN) eine wesentliche Informationsquelle dar, um eine ‚Innenperspektive‘ der FSP zu erhalten.

Ziel dieses Artikels ist daher eine erste Evaluation der ärztekammerlichen FSP aus der Sicht der TN, welche mithilfe einer *online* durchgeführten Umfrage ermittelt wurde. Dafür wird zunächst die aktuelle Situation ausländischer Ärzt\*innen in der Bundesrepublik Deutschland in den Blick genommen, wobei ein Unterkapitel (1.1) den gesetzlichen Weg zur Anerkennung im Ausland erworbener medizinischer Berufsqualifikationen zur Erteilung der Approbation bzw. Berufserlaubnis in der BRD darstellt. Die anschließend in Kapitel 2 angeführten Besonderheiten medizinischer Kommunikation und die Komplexität der Kompetenzen verdeutlichen den Umfang der Anforderungen dieses Berufsfelds und die damit verbundenen (sprachlichen) Hürden besonders für ausländische Ärzt\*innen. Im dritten Kapitel folgt die Vorstellung der Fachsprachenprüfung der Bezirksärztekammer Rheinhessen, deren Aufbau

exemplarisch für das im Bereich Medizin etablierte Prüfungsformat seitens der Ärztekammern herangezogen wird<sup>1</sup>. Der Aufbau der Online-Umfrage sowie die Perspektive der Prüfungsteilnehmenden werden in Kapitel 4 anhand der Umfrage-Ergebnisse präsentiert und abschließend diskutiert.

## 1 Zur Situation ausländischer Ärztinnen und Ärzte in Deutschland

Die Notwendigkeit eines gut funktionierenden Gesundheitssystems wurde durch die Corona-Pandemie verdeutlicht, aber auch sonst sind „Ärztinnen und Ärzte [...] systemrelevant“ (Ärztestatistik 2019), dies betonte auch der neue Bundesärztekammer-Präsident Dr. Klaus Reinhardt. Die derzeitige Arbeitsmarktsituation und aktuelle Prognosen weisen jedoch eine erhöhte Nachfrage u.a. bei Gesundheits- und Pflegeberufen auf (vgl. BfA 2019, Augurzky & Kolodziej 2018: 31). Darunter fallen auch Haus- und Fachärzt\*innen der Humanmedizin (vgl. ebd. und Abb. 1).

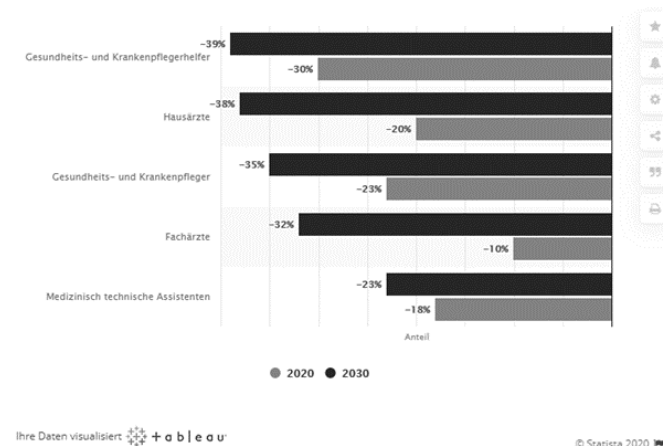


Abb. 1: Prognose zum Fachkräfteengpass in gesundheitswirtschaftlichen Berufen in Deutschland bis zu den Jahren 2020 und 2030 (Daten aus dem Jahr 2016, veröffentlicht 2020)

1 Vgl. auch Jehle in diesem Band.



Der immer wieder diskutierte bundesweite Fachkräftemangel (vgl. bspw. BfA 2019, Deppe & Borowski et al. 2015: 10) wird für diesen Berufsbereich kurz skizziert, da dieser letztlich den Grundstein der in diesem Beitrag thematisierten Fachsprachenprüfung darstellt.

In der Bundesrepublik Deutschland herrscht trotz der leicht steigenden bzw. gestiegenen Ärzt\*innenzahl ein Ärzt\*innenmangel, wie die Zahlen der jährlich erhobenen Studien der Bundesärztekammer belegen (vgl. Ärztestatistik 2019), da sich u.a. aufgrund der steigenden Lebenserwartung „die Schwere zwischen Behandlungsbedarf und Behandlungskapazitäten immer weiter [öffnet]“ (Montgomery 2017, zit. in Ärztestatistik 2016). Dies schafft einen Anreiz für ausländische Fachkräfte zum Arbeiten nach Deutschland zu kommen und birgt die Chance, den nicht gedeckten Bedarf mit qualifizierten Fachkräften aus dem Ausland zu kompensieren und diese Versorgungslücke gar zu schließen (vgl. BAMF 2015: 5, 21f.).

Zwar lassen sich für die Berufsgruppe der Ärzt\*innen weder eindeutige Push- und Pull-Faktoren nach dem gleichnamigen Modell von Everett S. Lee<sup>2</sup> anführen, noch gibt es aktuell statistische Erhebungen über die Zuwanderungsgründe von Mediziner\*innen. Doch zeigen die aktuellen Zahlen der Bundesärztekammer einen starken Zuwachs von Mediziner\*innen: In Deutschland waren im Jahr 2019 insgesamt 58.168 ausländische Ärzt\*innen gemeldet, was einen Anstieg um rund 3.800 zum Vorjahr bedeutet (+7,0%). Dabei kommt die größte Zahl aus Syrien (4.486), Rumänien (4.433), Griechenland (2.811), Österreich (2.381) und der Russischen Föderation (2.321) (vgl. Bundesärztekammer 2019: 36ff.). Insgesamt sind knapp 23.300 aller berufstätigen ausländischen Ärzt\*innen aus der EU, wenn man ganz Europa betrachtet, sind es rund 33.000.

.....

- 2 Das sog. ‚Push-Pull-Modell‘ beschreibt zwei Beweggründe für Migration: Menschen können von ihrem bisherigen Wohnort ‚weggedrückt werden‘ (englisch.: „to push“; deutsch: „drücken“) und bzw. oder von einem anderen Gebiet ‚angezogen‘ (englisch: „to pull“, deutsch: „ziehen“) werden (vgl. Winter-Heider 2009: 182). Demnach wird die Aussicht auf bessere Bildungschancen, soziale Sicherheit, eine attraktivere wirtschaftliche Situation oder auf bessere Gesundheitsversorgungssysteme als Pull-Faktoren definiert. Migrationsbewegungen aufgrund von Krieg, Gewalt, Verfolgung, Diskriminierung, wirtschaftlicher Armut, sozialpolitischen Problemen oder persönlicher Konflikte im Herkunftsland gelten, wie auch der Nachzug als Familienmitglied, als Push-Faktoren (vgl. ebd.: 183).

Dies ist nicht zuletzt auf die Richtlinien der Anerkennungsverfahren zurückzuführen, da hier zwischen EU- und Nicht-EU-Abschluss unterschieden wird.

Um den (neuen) Herausforderungen des deutschen Gesundheitssektors und der beschriebenen Personalnot entgegenzuwirken, kommt es auch zur bereits erwähnten gezielten Anwerbung ausländischer Fachkräfte, da „gerade in einer kurz- und mittelfristigen Perspektive [...] die Anwerbung von [Ärzt\*innen] aus dem Ausland als Option der Personalpolitik in Betracht gezogen werden [muss]“ (Hoesch 2003: 112). Dies betrifft ebenso den Bereich der Pflege: Im Jahr 2015 beschäftigten bereits 21,8 Prozent der deutschen Krankenhäuser Pflegekräfte aus dem Ausland (vgl. statista 2019). Dabei werden unterschiedliche Formen der Personalgewinnung genutzt: Neben Stipendienprogrammen für ausländische Medizinstudierende sowie approbierte Ärzt\*innen, Klinikmodellen zur konkreten Anwerbung, beauftragte Personaldienstleister\*innen bzw. -vermittler\*innen im In- und Ausland und Direktansprachen an Universitäten im Ausland (vgl. Adam & Croneberg 2012: 95, Vormweg 2012: 93), existiert auch ein ausgeweitetes Internetmarketing (vgl. Baumhove 2012: 96). So wirbt bspw. die private Sprachschule ‚Rheinland Institut‘ mittels ‚Online-Recruitment‘-Videos<sup>3</sup> mit der Jobvermittlung in einem deutschen Krankenhaus sowie der Unterstützung beim vollständigen Anerkennungsprozess (vgl. Rheinland Institut).

Das Anliegen seitens der Politik an einer verbesserten wirtschaftlichen Einbindung von Fachkräften mit Auslandsqualifikationen und die Integrationsförderung waren Ausgangspunkt gesetzlich geregelter Anerkennungsverfahren, die die Voraussetzung zur Arbeit als Ärztin bzw. Arzt in Deutschland festlegen. Diese werden im Folgenden erläutert.

## 1.1 Der Weg zur Approbation

Angesichts der oben dargestellten Situation stellt sich die Frage, wie medizinisch ausgebildetes Fachpersonal aus dem Ausland überhaupt in diesem Be-

.....

3 Vgl. ‚Doctor online recruiting in Germany‘, abrufbar unter: [https://www.youtube.com/watch?v=CaRiQBL-suOE&t=11s&fbclid=IwAR26LFmj2jqB7Zk8pfPQCNU-Na954ODH14UK-Ww4lhIgUkRmB\\_Vnk3bbIU](https://www.youtube.com/watch?v=CaRiQBL-suOE&t=11s&fbclid=IwAR26LFmj2jqB7Zk8pfPQCNU-Na954ODH14UK-Ww4lhIgUkRmB_Vnk3bbIU), 20.07.2020.

reich tätig werden kann bzw. wie Ärzt\*innen mit ausländischen Berufsqualifikationen die Approbation bzw. Berufserlaubnis<sup>4</sup> in Deutschland erhalten. Hierfür spielt sowohl die fachliche Anerkennung also auch der Nachweis fachspezifischer Sprachkenntnisse eine große Rolle. Für ersteres dient das im April 2012 in Kraft getretene *Gesetz über die Feststellung der Gleichwertigkeit von Berufsqualifikationen*, kurz *Berufsqualifikationsfeststellungsgesetz – BQFG* (vgl. bp-Portal) als Grundlage. Bei der Gleichwertigkeitsprüfung des Abschlusses wird danach unterschieden, ob dieser in der EU bzw. EWR oder in der Schweiz absolviert wurde oder in einem nicht-europäischen Ausland (s. Anhang). Ein nicht-europäischer Abschluss zieht nach Antragstellung bei der zuständigen Behörde des jeweiligen Bundeslandes einen deutlich längeren bürokratischen Prüfweg nach sich (vgl. Dippold-Schenk 2018: 34), wohingegen für die in der EU erworbenen Qualifikationen eine sofortige Erteilung der Approbation in Deutschland möglich ist (vgl. ebd.). Beabsichtigt man nach Erhalt der Approbation die Anerkennung seines Facharztes, folgt ein weiterer Antrag und eine Eignungs- oder Kenntnisprüfung bei der Landesärztekammer (vgl. Informationsportal der Bundesregierung).

Die Finanzierung der Antragsstellung und ggf. Anpassungsqualifizierungen ist kostspielig. Außerdem gibt es je nach Bundesland spezifische Umsetzungsregelungen, was eine komplette Vereinheitlichung erschwert. Dennoch lässt sich sagen, dass das in Krafttreten des BQFGs ein erster Erfolg auf dem Weg zur verbesserten und vereinfachten Anerkennung ausländischer Berufsqualifikationen darstellte (vgl. Braun 2012: 10). Darüber hinaus wurde für die Gewinnung qualifizierter Fachkräfte am internationalen Arbeitsmarkt im März 2020 das neue *Fachkräfteeinwanderungsgesetz* (FEG) erlassen, um so den Arbeitszugang aus Staaten außerhalb der EU zu erweitern und zu erleichtern (vgl. BMAS 2020).

Neben der Anerkennung fachlicher Qualifikationen sind wie bereits erwähnt fachspezifische Sprachkenntnisse zur Ausübung des jeweiligen Berufs unerlässlich. Anlässlich der 87. Gesundheitsministerkonferenz 2014 (GMK)

.....

4 Mit einer Berufserlaubnis kann im Krankenhaus oder einer Praxis gegen Bezahlung gearbeitet werden (max. zwei Jahre). Eine Weiterbildung zu einer Facharztqualifikation ist erst nach Erhalt der Approbation möglich (vgl. BÄK 2020a: 2).

wurde im Hinblick auf fachspezifische Sprachkenntnisse von akademischen Heilberufen Folgendes festgestellt:

In der Praxis hat sich jedoch gezeigt, dass die insoweit von Sprachinstituten angebotenen allgemeinsprachlichen Zertifikate und Diplome für die Überprüfung der für die Berufsausübung erforderlichen Sprachkenntnisse nicht geeignet sind (87. Gesundheitsministerkonferenz 2014: 2).

Ferner wurde im Beschluss dieser Konferenz festgehalten, dass von internationalen Ärzt\*innen neben allgemeinsprachlichen Sprachkenntnissen auf B2-Niveau (*GER*) auch fachspezifische Sprachkenntnisse auf C1-Niveau (*GER*) nachzuweisen sind (vgl. ebd.: 3).

Mit dem Inkrafttreten des obigen Beschlusses entstand der Bedarf an Überprüfungsmöglichkeiten der geforderten medizinischen Fachsprache und somit an einer Fachsprachenprüfung (FSP). Der Entwicklung von Deutsch-Prüfungen im Bereich Medizin nahm sich als Institution die *telc gGmbH* an, wobei die Durchführung in erster Linie an die Landesärztekammern in ‚Eigenregie‘ übertragen wurde. Hierzu ist vermerkt, dass „[d]ie Länder [...] dabei eine bundeseinheitliche Vorgehensweise an[streben]“ (87. GMK 2014: 7), was sich in der Praxis aber als schwierig erweist. Dies hängt u.a. mit der Komplexität der (Sprach)Kompetenzen und dem Umfang der Anforderungen an dieses Berufsfeld und letztlich auch mit den Anforderungen an eine solche FSP zusammen. Bevor auf die FSP eingegangen wird, erfolgt ein Überblick über die Besonderheiten der medizinischen Kommunikation.

## 2 Besonderheiten und Kompetenzen der medizinischen Kommunikation

Medizinische Kommunikation ist durch eine Vielfalt an Kommunikationsebenen, -formen und -situationen gekennzeichnet, an denen sowohl Ärzt\*innen und Patient\*innen als auch medizinisches Fach- bzw. Assistenzpersonal und Angehörige beteiligt sind (vgl. Roelcke 2005: 193). Außerdem kommt es im medizinischen Bereich, beispielsweise im Krankenhaus, oft zu fachdiszi-

plinübergreifender Zusammenarbeit, was eine Verständigung sowohl auf der berufs- als auch auf der fachsprachlichen Ebene zwischen Mediziner\*innen bedeutet. Je nach beteiligten Personen und Kontext muss die Ärztin/der Arzt also auch ihre/seine Sprache an die Situation anpassen. Dadurch entsteht „ein Spannungsfeld“ (Efing 2014: 420) zwischen der Lebens- bzw. Ausdruckswelt der Patient\*innen und einem professionellen Umgang mit medizinisch wissenschaftlichen Themen (vgl. Schrimpf & Bahnemann 2009: VI). Eine einheitliche Kommunikationsform von Ärzt\*innen lässt sich damit nicht definieren. Dies bedeutet, dass in Deutschland tätige ausländische Mediziner\*innen alle Kommunikationsebenen beherrschen müssen (vgl. ebd.), nicht zuletzt aus dem Grund, da „Kommunikationsfehler [...] im medizinischen Bereich – anders als im alltäglichen Leben – nicht nur ärgerlich, sondern [...] auch gefährlich“ sein können (Bechmann 2014: IX). „Mehr als 80% der medizinischen Fehlbehandlungen gehen nicht auf unzureichende Fähigkeiten, sondern auf mangelhafte Kommunikation zurück“ (Baller & Schaller 2017: 162), so ist das Patienten\*wohl und die Patienten\*sicherheit nachdrücklich durch Kommunikationserfolg bestimmt und dies nicht nur bei der Notfallbehandlung (vgl. ebd.: 7).

Darüber hinaus belegen wissenschaftliche Studien die Wichtigkeit ärztlicher Kommunikation, da ein signifikanter Zusammenhang zwischen dem Heilungsprozess und dessen Beeinflussung durch die medizinische Kommunikation bestehe (vgl. Ärztekammer Nordrhein 2015: 5). Das Gespräch zwischen Ärzt\*innen und Patient\*innen erfüllt folglich „vielfältige Funktionen im diagnostischen und therapeutischen Prozess“ (Langer & Schnell 2009: 7). Hinzu kommen kulturelle Unterschiede z.B. im Umgang mit Patient\*innen und Vorgesetzten, mit dem deutschen Versicherungs- und Verwaltungssystem, den Hierarchiestrukturen innerhalb von Krankenhäusern und allgemein mit dem deutschen Gesundheitswesen (vgl. Schrimpf & Bahnemann 2009: VII). Dies verdeutlicht die großen Herausforderungen, denen sich ausländische Mediziner\*innen auch sprachlich stellen müssen. Obwohl es schwer ist, klare Grenzen bei der Verwendung von Alltags-, Berufs- und Fachsprache zu ziehen und die Übergänge oftmals fließend sind (vgl. Efing 2014: 420), hat sich in der Praxis gezeigt, dass auf die Aneignung bzw. Vermittlung von fachspezifischen Sprachkenntnissen dennoch besonderer Wert gelegt werden muss (vgl. 87. GMK 2014: 2, Kapitel 2.1), da diese in der mündlichen wie schriftlichen Kommunikation u.a. spezifische „Mitteilungsstruk-

turen“ aufweisen (Buhlmann & Fearn 2000: 12). Dies äußert sich beispielsweise in der häufigen Verwendung von Abkürzungen und Wortableitungen, wie Reha für Rehabilitation oder EKG für das Elektrokardiogramm, von speziell im medizinischen Kontext immer wieder gebrauchten Redewendungen wie „Was machen die Schmerzen heute?“, „Ich taste nun ... ab“, „Liegt der Befund von ... vor?“ (Thommes & Schmidt 2016) oder im Formulieren fachlicher Informationen (z.B. in einem Arztbrief). Die verschiedenen Textsorten der medizinischen Fachkommunikation (z.B. Fallberichte, Arztbriefe, Handbücher) sind hierbei durch eine „ausgeprägte Makrostruktur“ (Roelcke 2005: 194) gekennzeichnet und dienen der/dem jeweiligen Adressat\*in zur „zuverlässigen Orientierung“ (ebd.).<sup>5</sup>

Die Sprachkompetenzen, die im Bereich Medizin notwendig sind, lassen sich in die verschiedenen Fertigkeiten Hören, Lesen, Schreiben und Sprechen unterteilen. Für eine einheitliche und nachvollziehbare Bewertung des Sprachniveaus gibt es nach dem *GER* sechs Niveaustufen. Gemäß dieser Einteilung sind für jede Fertigkeit und Niveaustufe allgemeine Kann-Beschreibungen formuliert worden, die festlegen, was man ‚können‘ muss, um ein gewisses Sprachniveau zu beherrschen (vgl. *GER*). Wie in Kapitel 1 bereits erwähnt, sind allgemeinsprachliche Kompetenzen für den medizinischen Bereich dennoch nicht ausreichend. So wurden spezielle Kann-Beschreibungen für Mediziner\*innen angefertigt. Dafür wurden die Stufenbeschreibungen der Globalskala des *GER* dem medizinischen Kontext angepasst und es wurde damit versucht, eine Antwort auf die Frage „Was können Ärztinnen und Ärzte auf den Kompetenzstufen B2 und C1?“<sup>6</sup> (*telc*) zu geben.

Die Ausführungen dort<sup>7</sup> zeigen die Komplexität der medizinischen (Fach-) Sprache und den vielseitigen Kommunikationseinsatz von Mediziner\*innen,

.....

- 5 Für eine ausführlichere Erläuterung der spezifischen Merkmale der medizinischen Fachsprache vgl. Roelcke 2005: 193ff.
- 6 Das vorgestellte Sprachniveau B2–C1 wurde bewusst gewählt, da es sich dabei um die Kompetenzstufen handelt, die zur Anerkennung im Ausland erworbener fachlichen Qualifikationen im medizinischen Bereich und zum Erhalt der Approbation in Deutschland vorausgesetzt wird (vgl. Kapitel 2.1).
- 7 Für eine zusammenfassende Darstellung vgl. Jehle 2019: 47f. Die vollständige Beschreibung der Kompetenzen von *telc* sind unter folgendem Link zu finden: [https://www.telc.net/file-admin/user\\_upload/Broschueren/ger\\_poster\\_medizin.pdf](https://www.telc.net/file-admin/user_upload/Broschueren/ger_poster_medizin.pdf), 21.07.2020.

dessen Beherrschung durch eine abzulegende FSP überprüft werden soll. Im Folgenden wird diese am Beispiel des Landes Rheinland-Pfalz skizziert.

### 3 Ärztliche Fachsprachenprüfung Rheinland-Pfalz

Die Konzeption und Durchführung der FSP erfolgt vorwiegend durch die Landes- bzw. Bezirksärztekammern des jeweiligen Bundeslandes. Ein einheitliches Prüfungssystem existiert bisweilen nicht<sup>8</sup> (vgl. Malakhova & Lenz 2018), dennoch bestehen aufgrund der Beschlüsse der 87. Gesundheitsministerkonferenz grundlegende Vorgaben zu den Inhalten. So sind mindestens ein Arzt\*-Patienten\*-Gespräch (Anamnese), das Verfassen eines Schriftstücks (Kurz-Arztbrief) und ein Arzt\*-Arzt\*-Gespräch (Patienten\*vorstellung) Bestandteile einer FSP (vgl. 87. GMK 2014: 6).

Die ärztliche Fachsprachenprüfung in Rheinland-Pfalz (RLP) wird im Auftrag der Landesärztekammer Rheinland-Pfalz und des Landesamtes für Soziales, Jugend und Versorgung (LSJV) in Koblenz von der Bezirksärztekammer Rheinhessen durchgeführt. Die dort durchgeführte Prüfung ist die einzige in RLP zur Beantragung der Approbation bzw. Berufserlaubnis anerkannte FSP (vgl. Marburger Bund 2018: 4). Der Wohnort bzw. der offizielle Wohnsitz oder das Nachweisen eines konkreten Stellenangebots ist Bedingung für das Ablegen der Prüfung in RLP bzw. im jeweiligen Bundesland. Die Anmeldung zur Prüfung ist nur nach Erhalt einer qualifizierten Eingangsbestätigung des LSJV in Koblenz im Zuge der Beantragung auf Approbation bzw. Berufserlaubnis möglich, außerdem sollte ein Sprachzertifikat vorliegen.<sup>9</sup> Das Niveau (nach GER) ist hierbei nicht spezifiziert.

Nach erfolgreicher Anmeldung erwartet die Prüfungsteilnehmenden eine 2,5–3 stündige Absolvierung (inkl. Wartezeiten) mit folgenden fünf bzw. sechs

.....  
8 Siehe dazu den Beitrag von Jehle in diesem Band.

9 Hier liegt eine deutliche Diskrepanz zwischen der Auflistung der für den FSP-Antrag benötigten Unterlagen auf der Informationsseite der Ärztekammer und der praktischen Handhabung vor, siehe dazu die Umfrageergebnisse in Kapitel 5.

Teilen und mehreren Abschnitten, die unterschiedliche Wertigkeit für die Gesamtprüfung haben<sup>10</sup>:

1. Verständnistest (früher Vokabeltest): Hier werden 20 medizinische Begriffe aus der Fachsprache ins Deutsche übersetzt und umgekehrt sowie eine Schemazeichnung des Körpers mit medizinischen Bezeichnungen beschriftet bzw. ausgewählte Körperteile (in „Laiensprache“) unterschiedlich markiert. Für die Vorbereitung dieser Aufgaben wird die Arbeit mit medizinischen Wörterbüchern empfohlen.
2. Freies Gespräch mit folgender Erläuterung: „Hier werden die Prüfer Ihnen Fragen stellen, um zu beurteilen, wie gut Sie Sprache verstehen und wie gut Sie Antworten formulieren können“.
3. Ein 20-minütiges Arzt\*-Patienten\*-Gespräch (Anamnesegespräch/Patienten\*aufklärung) in laienverständlicher Alltagssprache. Nicht eine vollständige Anamnese, sondern das Sprechen, Verstehen und Reagieren auf Patient\*innen steht hierbei im Vordergrund und führt bei Nichtbeachten der Aussagen und Fragen dieser zu erheblichem Punkteverlust.
4. Abfassen eines Schriftstücks (Kurz-Arztbrief): Hier soll eine aktuelle Anamnese in Form eines Fließtexts geschrieben und die erfragten Informationen aus dem vorangegangenen Anamnesegespräch stichwortartig erfasst werden. In Fachsprache folgen Ausführungen zu Verdachts- und Differenzialdiagnose(n) und zur zuvor mit der/dem Patient\*in mündlich besprochenen Untersuchungsforderung. Hierfür bekommen die Prüfungsteilnehmenden einen vierseitigen Anamnesebogen zum Ausfüllen, der auf der Internetseite zur

.....

10 Die Ausführungen beruhen auf den auf der Internetseite der Bezirksärztekammer veröffentlichten Angaben (vgl. BÄK 2020a). Trotz zusätzlicher ‚Ablaufhinweise zur FSP in Mainz‘ (vgl. BÄK 2020b) geben sie nicht umfassend Aufschluss über Details zu Prüfungsteilen, Bewertungskriterien und Prüfer\*innen. Auf meine Anfrage bei der Kammer diesbezüglich wurde leider nicht reagiert.



Vorbereitung einsehbar und herunterladbar ist<sup>11</sup>. Inkludiert ist das Ausfüllen von u.U. davon unabhängigen Anforderungsscheinen für weiterführende Untersuchungen.

5. Aufgaben zum Lese- und Hörverstehen sowie ein Arzt\*-Arzt\*-Gespräch (was nach Hinzufügen des freien Gesprächs eigentlich Teil 6 darstellen müsste<sup>12</sup>): Hier werden zu einem einseitigen Text (z.B. Befunde oder Arztbrief) drei Fragen auf einem Vordruck schriftlich beantwortet. Außerdem werden per Telefon (zwei Anrufe) patientenbezogene Informationen (z.B. aus dem Labor) übermittelt, die im Prüfungsbogen zu notieren sind. Bei dem Gespräch von Ärztin/Arzt zu Ärztin/Arzt handelt es sich um ein Fachgespräch, bei dem in berufstypischer Sprache ein Bericht über die/den Patient\*in aus Teil 1 oder die schriftlich bzw. akustisch übermittelten Informationen aus Teil 4 oder sonstige medizinische Themen erfolgt. An dieser Stelle wird betont, dass es nicht um medizinisches Fachwissen und dessen Überprüfung geht, sondern um „die Fähigkeit, sich gut auszudrücken und sich fachsprachlich miteinander verständigen zu können“. Medizinische Fehler oder fehlendes medizinisches Wissen werden nicht bewertet.

Die vorgestellten Prüfungsteile können hierbei in unterschiedlicher Abfolge geprüft werden (vgl. BÄK 2020a: 2). Dies übernehmen in der Regel zwei approbierte Ärzt\*innen, die auch die Rolle der Gesprächspartner\*in (Patient\*in und Kolleg\*in) einnehmen. Für das Bestehen muss jeder Prüfungsteil mit einer ihm zugeordneten Mindestpunktzahl abgeschlossen werden, wobei das Nichterreichen der erforderlichen Mindestpunkte in einem Teil zum sofortigen Abbruch der Prüfung führt. Die Prüfung gilt insgesamt als bestanden, wenn die Gesamtpunktzahl aus allen Prüfungsteilen bei über 60% liegt, d.h.

.....

- 11 Anamnesebogen verfügbar unter: <https://aerztekammer-mainz.de/pdf/diverse/FSP-Anamnesebogen-2020.pdf>, 11.08.2020.
- 12 Letzte Änderungen des Prüfungsablaufs bzw. der -teile erfolgten im Juni/Juli 2020. Zuvor: 1. Vokabeltest (ohne Körperschema), 2. Arzt\*-Patienten\*-Gespräch, 3. Verfassen eines Schriftstücks (Kurz-Arztbrief), 4. Lese- und Hör-Verstehen, 5. Arzt\*-Arzt\*-Gespräch. Gründe für die Ergänzungen sind nicht bekannt.

eine Addition der Mindestpunktzahlen aus allen Prüfungsteilen ist nicht ausreichend und muss durch höhere Punktzahlen in anderen Teilen ausgeglichen werden (vgl. ebd.). Das Prüfungsergebnis wird direkt im Anschluss an die Prüfung (mündlich) bekannt gegeben und bei bestandener Prüfung eine Bescheinigung (im Original) ausgestellt. Im Falle des Nichtbestehens kann die Prüfung wiederholt werden. Die Kosten für die Prüfung liegen derzeit bei 425€ (vgl. ebd.: 1).

Dies stellen die offiziellen und zugänglichen Angaben der FSP in RLP dar, über deren tatsächlichen Ablauf die TN einen spezifischeren Aufschluss geben können, weshalb in einer Studie genau diese Erfahrungen abgebildet werden sollten. Aufgrund der oft fehlenden Transparenz seitens der Ärztekammern ist das Ziel der Umfrage u.a. Einschätzungen zu den Prüfungsinhalten sowie dem -aufbau, dem Prüfer\*innenverhalten und der Bewertung zu erhalten. Ferner liefern die Aussagen der TN Anhaltspunkte für einen sich etablierenden Prüfungstourismus (vgl. Malakhova & Lenz 2018: 1), wonach die Wahl des Bundeslandes zum Ablegen der Prüfung nach dem vermeintlichen Schwierigkeitsgrad der FSP gewählt wird.

## 4 Pilotstudie

In Form einer quantitativen Online-Umfrage soll die ‚Innenperspektive‘ der Prüfung ermittelt werden und die Erfahrungen der Prüfungsteilnehmenden mit der von ihnen abgelegten FSP aufgezeigt werden. Die Zielgruppe der Umfrage waren demnach internationale Ärzt\*innen, die bundesweit bereits an einer FSP teilgenommen haben, das Ergebnis dieser Prüfung spielte hierfür keine Rolle. Die Beantwortung des digitalen Fragebogens erfolgte seitens der Umfrageteilnehmenden (TN) selbstständig und anonym.

Der Online-Fragebogen wurde mithilfe des Programms *LimeSurvey* erstellt und beinhaltet insgesamt 34 Fragen in verschiedenen Frageformaten wie Einfach-/Mehrfachauswahl von Antwortoptionen, Matrix-Auswahl, fünfstufige Likert-Skala sowie freie Texteingabe. Durch die Ergänzung nichtstandardisierter Fragen ist es trotz der meist vorgegebenen Antworten möglich, individuelle

Aspekte bzgl. bestimmter Sachverhalte zu ermitteln<sup>13</sup>. Die Fragen wurden in sechs Themenbereiche unterteilt: *Allgemeines zur Person und zur Prüfung*, zu den *Prüfungsinhalten* sowie dem *-aufbau*, zur *Vorbereitung* und *Sonstiges*. Alle Fragen wurden mit ‚Pflichtangabe‘ eingestellt. Durch die Rubrik ‚Sonstiges‘ wurde die Möglichkeit einer freien Antwort gegeben, sodass die Beantwortung der Fragen in jedem Fall gewährleistet werden konnte. Außerdem wurde bei den Einstellungen darauf geachtet, dass die TN innerhalb der Umfrage nicht zu den vorherigen Fragen zurückkehren können, um u.U. bereits abgegebene Antworten zu revidieren.

Der Online-Fragebogen bzw. dessen Verlinkung wurde v.a. über die Social Media Plattform *Facebook* gezielt in fachspezifische Gruppen in Umlauf gebracht. Zudem wurde bei einem Kursträger für vom BAMF-geförderte ‚Mediziner-Kurse‘ sowie einigen Beratungs- und Trainingsdiensten um die Weiterleitung an die Prüfungsteilnehmenden gebeten. Es handelte sich somit um eine Teilerhebung, die auf einer willkürlichen Stichprobenauswahl basiert. Von einer ‚zufälligen Auswahl‘ im methodischen Sinne kann insofern gesprochen werden, da keine persönliche Beeinflussung seitens der Umfrage-Durchführenden darauf erfolgte, wer aus der gewählten Zielgruppe letztlich an der Befragung teilnahm (vgl. Burzan 2015: 130).

Einige ausländische Ärzt\*innen nahmen dies zum Anlass eines direkten Austausches (per Mail) und berichteten über ihre individuelle Situation oder baten um Unterstützung in Bezug auf die FSP. Dies ermöglichte weitere Einblicke zu gewinnen.

## 4.1 Umfrageergebnisse

Die Online-Umfrage war bis zur Auswertung für drei Wochen aktiv<sup>14</sup>. Die hier präsentierten Umfrageergebnisse beziehen sich auf diesen Zeitraum. Zunächst wird die Stichprobe der teilgenommenen Ärzt\*innen kurz beschrieben

.....  
13 Für Ausführungen zur Fragebogenkonstruktion vgl. bspw. Schnell et. al. (2018): *Methoden der empirischen Sozialforschung*, Berlin/Boston: de Gruyter: 312ff.

14 Die Umfrage ist aufgrund des hohen Interesses und zwecks Validierung weiterhin online zugänglich unter: <https://survey.zdv.uni-mainz.de/index.php/395797?lang=de> (Stand Oktober 2020).

und anschließend die konkreten Ergebnisse der Umfrage dargestellt. Die Auswertung der durchgeführten Umfrage erfolgt anhand der zusammengestellten Antworten des Programms *LimeSurvey*. Statistiken und Interpretationen sind selbst erstellt<sup>15</sup>.

### *Allgemeines zur Person und zur Prüfung*

Im besagten Zeitraum von drei Wochen haben insgesamt 149 Personen an der Umfrage teilgenommen, wovon 59 Teilnehmende den Fragebogen vollständig ausgefüllt haben. Ein möglicher Grund mag der Umfang von 34 Items sein, ein spezifisches ‚Abbruchmuster‘ ist nicht erkennbar. Für die Analyse wurden 66 Fragebögen verwendet, da bei den sieben unvollständigen die letzten nicht beantworteten Fragen für diese Auswertung vernachlässigt werden können. Die TN-Gruppe bestand überwiegend aus Nicht-EU-Staatsbürger\*innen, die v.a. in Syrien, Ägypten (je 9 TN), im Irak (7 TN) und in der Ukraine (4 TN) ihre medizinische Ausbildung absolvierten. Dies spiegelt die Angaben des BAMF wider (vgl. Kapitel 1).

62 der 66 TN legten die FSP bei der jeweiligen Ärztekammer ab. Die Mehrheit absolvierte diese in Rheinland-Pfalz (RLP) (47), neben Baden-Württemberg, Bayern, Berlin, Brandenburg, Niedersachsen, Nordrhein-Westfalen, Sachsen und Thüringen. Dies ermöglicht, die Ergebnisse spezifischer für die Prüfung in RLP zu analysieren. Bei der Anerkennung zeigt sich Berlin besonders flexibel; dort wurden auch Prüfungen z.B. von *telc* und der PKT (Patientenkommunikationstest) anerkannt, obwohl es dort auch die ärztekammerliche FSP abzulegen gilt. Insgesamt haben etwa 70% der TN die Prüfung bestanden, 10 TN gaben zusätzlich an, nicht (gleich) bestanden zu haben, wobei sogar drei bis vier Versuche erfolglos waren. Die Bestehensquote in RLP unter den Befragten lag auch bei 70%, wobei als häufigste Gründe für das Nicht-Bestehen das Prüfungsformat und die Struktur (z.B. Markierungen der Körperteile und Vokabeltest) sowie die prüfende Kommission genannt

.....

15 Die Auswertung mehrerer Variablen bzw. Ergebnisse unterschiedlicher Fragen der Umfrage wurden in Korrelation zueinander gesetzt und mithilfe von Excel (grafisch) dargestellt. Auf eine genaue Ausführung der durchgeführten Analysetechnik wird im Rahmen dieses Beitrags verzichtet.

wurden. Unter denjenigen, die bewusst RLP als Prüfungsort gewählt hatten, ist immer wieder von einer kürzeren Wartezeit bei der Terminvergabe bzw. einem schnelleren Vorgehen insgesamt von der Anmeldung bis zum Ablegen der Prüfung die Rede. Nur ein/e TN gab an, dass sie/er die Prüfung anstatt in Mainz lieber in Rostock hätte ablegen wollen, da diese angeblich „viel einfacher“ sei und „weniger Fachbegriffe“ (ID 98, F7)<sup>16</sup> abgefragt würden.

Von den 47 TN, die die FSP bestanden haben, sind 27 TN in Besitz einer Approbation und haben in Deutschland eine Arbeitsstelle. Die niedrige Zahl lässt sich sicherlich mit den langen Wartezeiten auf die bei Nicht-EU-Bürger\*innen erforderliche Kenntnisprüfung erklären (vgl. Kapitel 1.1). Sieben FSP-Absolvent\*innen sind auch ohne Approbation bereits als Ärztin/Arzt tätig (in Berlin, RLP, Bayern, Thüringen), was im Rahmen einer Berufserlaubnis möglich ist (vgl. Kapitel 1.1). In Bezug auf den Arbeitsort konnte festgestellt werden, dass über 70% der TN, die in RLP die FSP abgelegt haben und als Ärzt\*innen tätig sind, in einem anderen Bundesland arbeiten und als ein Anzeichen für den vermeintlichen Prüfungstourismus gedeutet werden kann. Sonst entspricht der Arbeitsort auch dem Bundesland der abgelegten Prüfung.

Den Prüfungsablauf empfanden insgesamt rund 65% als transparent (RLP: 63%), wobei mehr als ein Viertel (RLP: 34%) angab, dass die Prüfung nicht oder nur teilweise wie erwartet durchgeführt wurde. Der Umfang wurde bei der am häufigsten durchgeführten ‚3-Teile-Prüfung‘ (vgl. Kapitel 3) mehrheitlich mit „genau angemessen“ eingestuft, der Schwierigkeitsgrad allerdings von mehr als der Hälfte als „schwer“ oder sogar „zu schwer“.

Zu den Prüfer\*innen machten die TN folgende Angaben: Die Prüfenden schnitten in den Kategorien „gut vorbereitet“ und „fachlich kompetent“ am besten ab, wohingegen sie nicht als besonders „fair“ und „motiviert“ sowie noch weniger als „geduldig“ galten. Hierbei decken sich die allgemeinen Einschätzungen mit denen in RLP. Trotz des ausdrücklichen Verweises im GMK-Beschluss (87. GMK 2014: 6) als auch im Informationsblatt der Be-

.....

16 Dies stellt die Zuordnung der Umfrageergebnisse dar. ID 98: Teilnehmendenummer, anonymisiert; F7: „Frage 7“, Bezifferung der Fragen nach Erstellung im Programm LimeSurvey. Die Reihenfolge des Erstellens und die Nummerierung sind für die Auswertung nicht von Bedeutung.

zirksärztekammer Rheinhessen für die Prüfung in RLP, dass das Fachwissen der Ärzt\*innen weder getestet noch bewertet wird, wurde bei 65% der TN (RLP: knapp 60%) auch medizinisches Fachwissen abgefragt. Wenn man bedenkt, dass es sich um eine Sprachprüfung handelt und nicht alle TN eine Facharztausbildung absolviert haben (29% der Befragten), führt dies unweigerlich zur Benachteiligung und u.U. sogar zur ‚Verfälschung‘ der Bewertung. Ein/e TN schreibt hierzu: „Es war nicht klar, wie ich genau antworten sollte entweder ist es über medizinische Kenntnisse orientiert oder Kommunikation und Sprachfähigkeit“ (ID 79, F10).

Durchgängig werden bei mehreren Fragen im Hinblick auf den Prüfungsablauf, die Bewertung (skriterien) und die allgemeine Zufriedenheit immer die Prüfer\*innen bzw. die Prüfungskommission und deren Verhalten erwähnt, was u.a. als „subjektiv“, „unfair“, „unangenehm“ und sogar „rassistisch“ beschrieben wird sowie das Bestehen stark von ihnen abhängen würde. Trotzdem werden im Allgemeinen die Bewertungskriterien von 65–71% der TN als „fair“, „transparent“ und „angemessen“ gesehen, in RLP fällt diese Beurteilung schlechter aus: „fair“ und „transparent“ finden diese um die 60%, „angemessen“ jedoch nur 27%. Ein/e Befragte/r äußert sich zu diesem Thema wie folgt:

*Ich habe mich für diese Prüfung vollständig vorbereitet. Alle Details aus verschiedenen Büchern habe ich gelernt. Auf der anderen Seite war das Niveau der Prüfung echt niedrig. Ich kann mir gut vorstellen, dass viele ausländische Kollegen besonders aus Drittstaat die Prüfung bestehen können, ohne notwendige Fertigkeiten zur Arbeit als Arzt in Deutschland zu haben. Das ist echt schlecht. Höhe bestandenquote in dieser Prüfung ist peinlich für Deutschlanda Gesundheitswesen sowie Kliniken, die diese Kollegen mit nicht genug Deutschkenntnissen einstellen werden [sic] (ID 120, F11).<sup>17</sup>*

Des Weiteren bestätigten die Befragten, dass in den meisten durchgeführten FSP eine/einer der prüfenden Ärztinnen/Ärzte auch die Rolle der/des Ge-

.....

17 Die Aussagen der Umfrage-Teilnehmenden wurden im Original übernommen und bewusst nicht korrigiert.

sprachspartner\*in beim Anamnese-Gespräch und bei der Patienten\*vorstellung einnahm. Dies wird unterschiedlich kommentiert, zum einen nehmen TN an, Ärzt\*innen könnten besser Symptome erklären und die „Arbeit“ besser beurteilen, zur FSP in RLP heißt es jedoch bspw. auch:

*Die Prüfung in Mainz ist sehr subjektiv. [xy]<sup>18</sup> war der Prüfer und meint, dass meine Aussprache zum [Miss]Verständnis führe. Dann habe ich B2.C1 telch FSP abgelegt und bestanden. Meiner Meinung nach müssen bei einer FSP ein Arzt/in und ein Germanist/in als prüfer sein [sic]. (ID 67, F6)*

Die große Mehrheit der TN absolvierte eine FSP mit den üblichen drei Teilen, wie im GMK-Beschluss beschrieben, bei denen ein einziger Patienten\*-Fall behandelt wurde. Erstaunlicherweise gaben 11 TN an, in RLP bei der FSP nicht mit einem ‚Ein-Fall-Szenario‘ gearbeitet zu haben. Dies widerspricht jedoch jeglichen Informationen seitens der Ärztekammer, Kolleg\*innen und weiteren Prüfungsteilnehmenden. Die Frage, ob unterschiedliche Patienten\*-Fälle für die einzelnen Teile bevorzugt würden, wurde klar mit „Nein“ (85%) beantwortet.

### *Prüfungsinhalte und -aufbau*

Es wurde nach dem zur Verfügung stehenden Zeitrahmen gefragt. Die je 20 Minuten pro Teil wurden von fast allen als entweder „genau richtig“ (50%) oder sogar „zu kurz“ (44%) eingeschätzt. Die im Berufsalltag knapp abrufbare Zeit würde diese TN sicherlich vor große Herausforderungen und zeitliche Hürden stellen.

Die Prüfungsinhalte wurden insgesamt von mehr als 60% als „sehr relevant“ oder „relevant“ und von knapp 40% als „teilweise relevant“ oder „nicht relevant“ für den Berufsalltag eingestuft. Die Verteilung verhält sich ähnlich in

.....

18 Ein Arzt der Prüfungskommission in Mainz wird mehrmals von verschiedenen Befragten namentlich genannt und scheint unter den Prüfungsteilnehmenden bekannt und ‚berühmt‘ für seine ‚Strenge‘ zu sein. Diese wird seitens einer Kollegin, die ihn bereits bei einer Hospitation bei der FSP erlebte, eher als ‚genau‘ bezeichnet, der v.a. auch sprachliche Nuancen wie Humor und Ironie mitbewertet.

RLP, wobei hier hauptsächlich die enorme Anzahl an medizinischen Wörtern, die sowohl in Fach- als auch in Laiensprache für den Vokabeltest (auswendig) gelernt werden, für irrelevant angesehen wird. Zudem wünschen sich einige TN spezifischere Themen/Fälle, die die Fachrichtung der Kandidat\*innen berücksichtigen.

In diesem Zusammenhang wurden die TN außerdem gefragt, was sie außerhalb der fachlichen Kommunikationsebenen als prüfungsrelevant hielten. Dabei wählten die TN vorwiegend „ein Aufklärungsgespräch vor einer Operation“ und „den Austausch mit dem Pflegepersonal“. Außerdem wurde mehrfach die Aufklärung von weiterführenden Untersuchungen (z.B. MRT, CT) genannt, was das Training und Überprüfen der ‚Patienten\*aufklärung‘ bekräftigt. Ein/e Befragte/r bemängelte, dass in der FSP nicht genügend Kommunikationsfähigkeiten geprüft würden. Im Hinblick auf die Prüfungsform würden sich dennoch nur 11 der TN für eine um die genannten Kommunikationssituationen erweiterte Variante entscheiden. Diejenigen, die sich für die übliche Form mit drei Teilen bzw. Kommunikationssituationen aussprachen, nannten dafür bspw. den damit verbundenen erhöhten Lernaufwand und Stress, dass diese Teile das Minimum an ärztlichen Kenntnissen darstelle und diese zur Sprachbeurteilung ausreichen würden. Außerdem wäre bei weniger Teilen eine Vertiefung möglich und bei mehreren kürzeren bestünde das Risiko für Redundanz. Ein/e TN ergänzte:

*Aufklärungen dürfen wir im Alltag mit ein Aufklärungsbogen durchführen, muss nicht auswendig gemacht werden, deswegen sehe ich kein Grund dass das ein Teil von der Prüfung sein soll [sic]. (ID 106, F23)*

Diese Rückmeldungen spiegeln eine verbreitete Orientierung des reinen ‚Prüfungs-Bestehens‘ wider statt einer notwendigen ‚Kompetenzorientierung‘, besonders hinsichtlich der Relevanz für das Berufsleben. Für 30% der TN ist die Prüfungsform oder die Anzahl der Teile unwichtig, solange es eine gerechte und faire Prüfung ist, die letztlich bewertet, wie gut man vorbereitet ist bzw. wie gut man Deutsch sprechen und schreiben kann. Ein/e TN erwähnt in diesem Zusammenhang die Problematik des Auswendiglernens:



*[...] Der Inhalt der Prüfung muss unerwartet sein, sodass Kolleginnen und Kollegen nicht bestimmte Sätze lernen können [sic]. (ID 120, F24)*

Bislang finden die FSP in Präsenz statt. Bei der Umfrage wurden die TN zur ihrer Meinung über eine digitale Version der FSP befragt. 28 TN sprachen sich dafür und 32 eher dagegen aus. Bei einem digitalen Angebot wählten etwas weniger „bereits im Heimatland“ statt „in Deutschland“. Als häufigste Begründung, eine digitale Prüfung in der Heimat ablegen zu wollen, wurde die Zeit- und Kostenersparnis genannt. Weniger Stress und Nervosität werden diesem Format auch zugeschrieben. Allerdings drückt diesbezüglich ein/e Befragte/r folgende Bedenken aus:

*Online Prüfungen kommen mit großer Gefahr von einem Betrug. Leider Manche Menschen sind werden mit gefälschten Unterlagen in Deutschland als Arzt anerkannt. Ich habe viele gesehen ohne ärztliche Kenntnisse besonders aus arabischem Raum. Sie können keine EKG oder Laborbefunde interpretieren. Selbstverständlich, dass deutsche Patienten in besonderer Gefahr stehen werden [sic]. (ID 120, F31)*

### *Vorbereitung*

Insgesamt wird die Prüfungslandschaft bzw. das Anerkennungsverfahren von mehr als 70% als „(sehr) unübersichtlich und komplex“ bewertet. Als Informations- und Beratungsquellen dienen den TN hauptsächlich das Internet sowie Freunde und Bekannte mit FSP-Erfahrung, Beratungsstellen in Deutschland wurden auch vermehrt genutzt.

Der Großteil der Befragten (ca. 65%) gab an, sich im Selbststudium sprachlich auf die FSP vorbereitet zu haben, gefolgt von privaten Sprachkursen zur Prüfungsvorbereitung (ca. 38%) und Online-Training (ca. 32%). Lediglich 11 TN nahmen an den BAMF-geförderten Kursen „Spezialmodul Heilberufe“ teil. Neben einer Vorbereitung auf das Prüfungsformat gilt ein allgemeines Sprachniveau auf B2-Niveau durch die 87. GMK als vorgeschrieben und ist normalerweise bei der Beantragung zur Approbation bzw. zur Anmeldung der FSP vorzulegen. Die meisten der Befragten waren in Besitz eines anerkannten B2-Zertifikats, 13 TN gaben an, dass sie über ihre B2-Kenntnisse lediglich

einen Nachweis einer Sprachschule hatten. Und zwei Personen, die die FSP in RLP ablegten, konnten sogar kein B2-Niveau nachweisen. Wie in Kapitel 3 erwähnt, ist laut des Informationsblatts der Ärztekammer in RLP dem Antrag zur FSP eine Kopie eines Sprachzertifikats beizufügen, wobei eine Nennung des Sprachniveaus fehlt (vgl. BÄK Rheinhessen 2020: 1) und die FSP letztlich auch ohne Einreichen eines Sprachzertifikats abgelegt werden kann.

### *Sonstiges*

Die vielen sehr fehlerhaften ‚freien Kommentare‘ seitens der TN bekräftigen den Eindruck mangelhafter Sprachkenntnisse und veranschaulichen eine deutliche Diskrepanz: ‚Sprachliche Defizite‘ gelten laut TN einerseits oft als Grund für das Nichtbestehen der FSP seitens der Prüfungskommission, andererseits werden TN ohne einen B2-Nachweis offensichtlich zur FSP zugelassen bzw. bestehen diese sogar ohne allgemeine Sprachkenntnisse auf B2-Niveau. Nach Aussagen eines TN ist dies auch der Grund, warum die FSP in RLP als „einfachere“ gewählt würde. Er habe die FSP beim zweiten Mal trotz nicht vorhandener B2-Sprachkenntnisse und ohne Nachweis allgemeiner Sprachkenntnisse bestanden. Dass es sein Zweitversuch war, habe ihm sicherlich dabei geholfen. Das Zulassen ohne erforderliches B2-Zertifikat und folglich erhöhter Durchfallquote wirft die Frage nach einer wirtschaftlich motivierten Durchführung der FSP auf. Dazu berichtet ein/e TN passend:

*Ich hab von Anfang wusste, dass ich die Prüfung nicht bestehe. Der Sekretär wurde von jemandem gefragt, wie viel Leute haben wir heute und wie viel von ihnen sind im ersten Versuch. Leider war ich der einzige im ersten Versuch (und hab nicht bestanden). Ich glaube es geht nur um Geld und gibt keine logische Grund dafür jemand durchfallen zu lassen. Ich rede in meinem Fall [sic]. (ID 54, F10)*

Und fügt später hinzu:

*Die Skruktur der Prüfung muss sich ändern. Es geht um einer Arbeit im Krankenhaus. Und dort wird du mehrmals etwas kontrollieren bevor sie*

*antreten oder machen. Deswegen die Form der Prüfungen heutzutage richtet sich nur nach Finanzielle Profis und nichts mehr [sic]. (ID 54, F31)*

Als weitere (Änderungs)Vorschläge für die Prüfung und zur Verbesserung der Sprachkenntnisse können u.a. folgende Äußerungen seitens der Befragten genannt werden:

*Manche Kollegen lernen nur bestimmte Fälle und memorisieren bestimmte Sätze, um die Prüfung bestehen zu können. Der Inhalt der Prüfung muss variant und vielfältiger sein, sodass inkompetente Kollegen sich mehr für lernen engagieren [sic]. (ID 120, F22)*

*Ich denke, es wäre besser, wenn die Ärztekammer und die Krankenhäuser kooperieren, damit die Ärzte eine Hospitation vor der Prüfung und Arbeit könnten. Meiner Meinung nach wäre das sehr hilfreich, die Sprche der Ärzte/innen zu verbessern [sic]. (ID 41, F31)*

Die vorgestellten Ergebnisse geben nur einen sehr skizzenhaften Einblick darüber, wie die zukünftigen Ärzt\*innen die FSP empfinden, aber gerade die freien Kommentare sind schon sehr aufschlussreich und bekräftigen, das Prüfungsformat noch weiter zu untersuchen.

## 5 Fazit und Ausblick

Im Rahmen dieser Umfrage wurden die relevantesten Ergebnisse und somit eine erste Evaluation der FSP seitens der TN präsentiert, was v.a. eine ‚Innenansicht‘ der vorgestellten FSP in RLP ermöglichte. Das Ziel der Repräsentativität und der Möglichkeit einer Quantifizierung der (Umfrage- und Analyse-) Ergebnisse auf die Gesamtheit der Prüfungen kann mithilfe dieser Erhebung nicht zufriedenstellend realisiert werden. Die Ergebnisse ermöglichen jedoch einen Einblick auf die aktuelle Prüfungssituation sowie in die Erfahrungswelt der Prüfungsteilnehmenden und lassen einen Rückschluss auf die Anforderungen der FSP als Nachweis medizinischer Fachsprachenkompetenzen zu.

Das vorgestellte Prüfungsformat stellt, nach Vorgaben des GMK-Beschlusses, eine Fachsprachenprüfung dar, die medizinspezifische Sprachkenntnisse (auf C1-Niveau) überprüft. Hinreichend valide scheint die FSP jedoch nicht zu sein, da die Umfrage ein vermehrtes Überprüfen medizinischen Fachwissens ergab und laut einigen TN nicht ausreichende Kommunikationsfähigkeiten und für den Beruf unzureichende Deutschkenntnisse geprüft würden. Die Frage ‚Was genau gemessen werden soll?‘ sollte also hinsichtlich der in Kapitel 2 dargelegten Kommunikationskompetenzen zur Ausübung des ärztlichen Berufs erneut gestellt werden bzw. diese auch transparenter gemacht werden.

Darüber hinaus ist bereits die Zulassung zur FSP und sogar das Bestehen ohne (Zertifikats-) Nachweis allgemeiner Deutschkenntnisse auf B2-Niveau, was aus Sicht der TN eher als Vorteil empfunden wird, bedenklich. Berufe, bei denen die Patienten\*sicherheit oberstes Gebot darstellt und die Gesundheit auf dem Spiel stehen kann, bergen durchaus beachtliche Risiken. Der Personalmangel im medizinischen Bereich sollte und darf nie zu einem Herabsetzen der zur Berufsausübung fachlichen wie sprachlichen Anforderungen führen, auch nicht zu Berufsbeginn. Die Einschätzung, dass die durch die FSP nachgewiesenen Sprachkenntnisse zu niedrig für das souveräne Bestreiten des Berufsalltags sind, teilen auch andere zugewanderte Mediziner\*innen, wie in einer Studie von Borowski gezeigt wurde (vgl. 2018: 359). Außerdem sprach sich das deutsche<sup>19</sup> Klinikpersonal bei einer Befragung durch Soppa mehrfach für „ein besseres Sprachniveau [...] als Lösungsvorschlag [aus], um als ausländischer Arzt oder Ärztin an einer deutschen Klinik Fuß fassen zu können“ (2019: 55). Außerdem zeigte sich bei dieser Erhebung, dass „Patienten/innen [...] sprachliche Fähigkeiten und Fachkompetenz eng miteinander zu verknüpfen [scheinen]. Wenn sie ihre Ärzte/innen nicht gut verstehen, fühlen sie [sich] nicht adäquat behandelt, was in Verunsicherung münden kann“ (ebd.: 54), nicht zuletzt deswegen sind ein hochangesetztes Sprachniveau als auch ‚strenge‘ Bewertungskriterien essenziell.

Eine Einheitlichkeit der FSP in den Bundesländern bei Durchführung und Bewertung wäre hierfür sehr förderlich. Grundlegend ist darüber hinaus die

.....

19 Unter ‚deutsch‘ werden hier Personen bezeichnet, deren Muttersprache Deutsch ist sowie in Deutschland sozialisiert wurden.

Einstellung der internationalen Ärzt\*innen, die sich nicht ausschließlich auf ihre fachlichen Kenntnisse und medizinischen Fähigkeiten verlassen dürfen, sondern der Sprache und ihrer kompetenten Verwendung (auf allgemein- und fachsprachlicher Ebene) höhere Wichtigkeit beimessen müssen. Dies ist letztlich auch für das Bestehen der FSP ausschlaggebend.

Mögliche Zusammenhangshypothesen wie z.B. ein Bezug zwischen einer Bestehensquote und der Länge des Aufenthalts in Deutschland vor Ablegen der FSP, welche Rolle die Form des allgemeinen Deutscherwerbs als auch die Prüfungsvorbereitung sowie das Herkunfts- bzw. Ausbildungsland spielt, müssten über diesen Beitrag hinaus untersucht werden. Außerdem wären weitere Erhebungen zu den durch die Rückmeldungen entstandenen Fragen auch im Vergleich zu anderen Bundesländern interessant und könnten vertiefende Einblicke verschaffen. Berlin bspw. ist in der Umfrage bei drei Punkten positiv aufgefallen: einerseits die in Kapitel 4 beschriebene Flexibilität bei der Anerkennung anderer Prüfungen, andererseits werden dort bei der ärzttekammerlichen FSP laut der TN bei der Anamnese und bei der Patienten\*vorstellung vermehrt externe Personen als Gesprächspartner\*in eingesetzt. Außerdem sind die Bewertungskriterien für die drei Teile transparenter. Musterbewertungsbogen werden dafür *online* von der Ärztekammer bereitgestellt<sup>20</sup>. Von den wenigsten Ärztekammern, wie auch in RLP, werden das methodische Vorgehen also auch die Bewertungskriterien offengelegt. Die Objektivität der Bewertung wird zudem durch die Rollenübernahme der Gesprächspartner\*innen seitens der prüfenden Ärzt\*innen vermindert und es wird in der Umfrage vermehrt eine Abhängigkeit der Testergebnisse von den jeweiligen Prüfenden beschrieben.

Der Transparenz allgemein und einer prüfungsorientierten Vorbereitung würden Hospitationsmöglichkeiten bei der FSP z.B. seitens der unterrichtenden KL zugutekommen. Ferner könnten mit einer Ausweitung des Kursangebots v.a. der BAMF-geförderten Maßnahmen, um die durchaus hohen Kosten für Kurse, Prüfungen und Approbationsverfahren zu minimieren, mehr ausländische Ärzt\*innen erreicht und damit das Training wichtiger Kommunika-

.....  
20 Bewertungskriterien bzw. die Musterbewertungsbogen der Ärztekammer Berlin, abrufbar unter: [https://www.aerztekammer-berlin.de/10arzt/61\\_Fachsprachpruefung/index.html](https://www.aerztekammer-berlin.de/10arzt/61_Fachsprachpruefung/index.html), 19.08.2020.

tionskompetenzen gefördert werden. Eine gezielte Vorbereitung erhöht auch die Chancen, die FSP bereits beim ersten Durchlauf zu bestehen.

Bei der Umfragerhebung konnte ein hohes Mitteilungsbedürfnis seitens der Prüfungsteilnehmenden festgestellt werden, was sicherlich Anlass für weitere Forschungen gibt. Der Fokus könnte dabei auf die Prüfer\*innen, deren Schulung und Verhalten sowie auf die Bewertungskriterien gelegt werden, um somit auch die TN-Rückmeldungen besser beurteilen zu können. Ein Austausch mit den prüfenden Ärzt\*innen könnte hierbei Aufschluss über ‚die andere Seite‘ geben und eine Stellungnahme zu den oft sehr kritischen Äußerungen der TN gegenüber der Prüfungskommission ermöglichen.

Ferner betont Roche in diesem Zusammenhang, dass „eine voreilige Diagnose von sprachlichen Fertigkeiten ohne die wissenschaftliche Fundierung solcher Tests [...] verhängnisvoll werden [kann]“ (2013: 108). Der neue, von ihm an der LMU und dem Klinikum der Universität München 2018 mitentwickelte *Sprachtest für ausländische Mediziner*<sup>21</sup> (kurz SAM) setzt genau hier an und beruht für eine zuverlässige Bewertung fachspezifischer Sprachkompetenzen auf einem wissenschaftlich fundierten Konzept. Ob die Etablierung des SAM eine Lösung der dargestellten, durchaus problematischen Prüfungssituation sein kann, muss genauer betrachtet und vergleichend analysiert werden<sup>22</sup>.

## Bibliografie

- ADAM, U. & CRONENBERG, C.-H. (2012): „Akquisition und Einsatz ausländischer Ärzte im Studium, in der Hospitation und in der Weiterbildung – Katholische Kliniken im Kreis Kleve Trägerschaft mbH“, in: KLEVEN KIRCHNER H., SCHROETER M. & FLESCHE, M. (Hrsg.): *Personalakquise im Krankenhaus. Ärzte gewinnen, binden, entwickeln*, Berlin/Heidelberg: Springer: 94–96.
- AUGURZKY, B. & KOŁODZIEJ, I. (2018): *Fachkräftebedarf im Gesundheits- und Sozialwesen 2030. Gutachten im Auftrag des Sachverständigenrates*

.....  
21 Im Original ungegendert.

22 Siehe dazu den Beitrag von Jehle in diesem Band.

- zur Begutachtung der gesamtwirtschaftlichen Entwicklung, RWI – Leibniz-Institut für Wirtschaftsforschung, abrufbar unter: [https://www.sachverstaendigenrat-wirtschaft.de/fileadmin/dateiablage/gutachten/jg201819/arbeitspapiere/Arbeitspapier\\_06-2018.pdf](https://www.sachverstaendigenrat-wirtschaft.de/fileadmin/dateiablage/gutachten/jg201819/arbeitspapiere/Arbeitspapier_06-2018.pdf), 25.07.2020.
- BALLER, G. & SCHALLER, B. (2017): *Kommunikation im Krankenhaus. Erfolgreich kommunizieren mit Patienten, Arztkollegen und Klinikpersonal*, Berlin/Heidelberg: Springer.
- BAUMHOVE, O. (2012): „Keine Chancen ohne Web“, in: KIRCHNER H., SCHROETER, M. & FLESCHE, M. (Hrsg.): *Personalakquise im Krankenhaus. Ärzte gewinnen, binden, entwickeln*, Berlin/Heidelberg: Springer: 96–97.
- BECHMANN, S. (2014): *Medizinische Kommunikation. Grundlagen der ärztlichen Gesprächsführung*, Tübingen: Narr.
- BOROWSKI, D. (2018): *Sprachliche Herausforderungen ausländischer Anästhesist(inn)en bei Aufklärungsgesprächen. Eine gesprächsanalytische Studie zu „Deutsch als Zweitsprache im Beruf“*, Berlin: Frank & Timme.
- BRAUN, D. (2012): „Einheitlicher, transparenter, effektiver? Das Verfahren zur Bewertung von im Ausland erworbenen Qualifikationen im Wandel“, in: *Focus Migration 18*, Kurzdossier, abrufbar unter: [https://m.bpb.de/system/files/dokument\\_pdf/Kurzdossier\\_Qualifikationsanerkennung\\_2012pdf\\_0.pdf](https://m.bpb.de/system/files/dokument_pdf/Kurzdossier_Qualifikationsanerkennung_2012pdf_0.pdf), 20.07.2020.
- BUHLMANN, R. & FEARNES, A. (2000): *Handbuch des Fachsprachenunterrichts*, Tübingen: Narr.
- BURZAN, N. (2015): *Quantitative Methoden kompakt*, Konstanz/München: UVK.
- DEPPE, J. et al. (2015): *Praxisguide Deutsch im Krankenhaus. Die sprachliche Integration internationaler Ärztinnen und Ärzte. Eine Handreichung für Klinikleitungen, Deutschlehrkräfte und Bildungsfachleute*, Hamburg: Fachstelle Berufsbezogenes Deutsch im Förderprogramm IQ, abrufbar unter: [http://www.deutsch-am-arbeitsplatz.de/fileadmin/user\\_upload/PDF/Praxisguide\\_Web.pdf](http://www.deutsch-am-arbeitsplatz.de/fileadmin/user_upload/PDF/Praxisguide_Web.pdf), 27.07.2020.
- DIEKMANN, A. (2009): *Empirische Sozialforschung. Grundlagen Methoden Anwendungen*, Hamburg: Rowohlt Taschenbuch Verlag.
- DIPPOLD-SCHENK, K. (2018): *Die Entwicklung eines didaktischen Designs im Rahmen einer modularen Weiterbildungsreihe zur Professionalisierung*

- von Sprachlehrenden für den Bereich Deutsch-Medizin. Dissertation, abrufbar unter: <https://opus4.kobv.de/opus4-fau/frontdoor/index/index/year/2018/docId/9735>, 20.07.2020
- EFING, C. (2014): „Berufssprache & Co.: Berufsrelevante Register in der Fremdsprache. Ein varietätenlinguistischer Zugang zum berufsbezogenen DaF-Unterricht“, in: *Info DaF 4*: 415–441, abrufbar unter: [http://www.daf.de/downloads/InfoDaF\\_2014\\_Heft\\_4.pdf#page=3&view=Fit](http://www.daf.de/downloads/InfoDaF_2014_Heft_4.pdf#page=3&view=Fit), 26.07.2020.
- HOESCH, K. (2003): „Green Card‘ für Ärzte? Von der ‚Ärztenschwemme‘ zum Ärztemangel im deutschen Gesundheitssektor“, in: HUNGER, U. & KOLB, H. (Hrsg.): *Die deutsche Green Card. Migration von Hochqualifizierten in theoretischer und empirischer Perspektive*, Osnabrück: Universität Osnabrück (IMIS-Beiträge 22): 99–113.
- JEHLE, N. (2019): „Vermittlung fachspezifischer Fremdsprachenkompetenzen am Beispiel des Pilotprojekts ‚Fachspezifischer Deutschkurs: Humanmedizin‘ für Erasmus-Studierende an der JGU Mainz“, in: ERSCH, C. M. (Hrsg.): *Kompetenzen in DaF/DaZ*, Berlin: Frank & Timme: 41–66, abrufbar unter: <https://library.oapen.org/handle/20.500.12657/24334>, 21.10.2020.
- LANGER, T. & SCHNELL, M. W. (2009): „Einführung“, in: DERS. (Hrsg.): *Das Arzt-Patient-/Patient-Arzt-Gespräch*, München: Hans Marseille Verlag: 7–13.
- MALAKHOVA, O. & LENZ, H. (2018): *Fachsprachprüfung Medizin in Deutschland. Der Versuch einer ersten Bestandsaufnahme*. Unveröffentlicht.
- ROCHE, J. (2013): *Fremdspracherwerb. Fremdsprachendidaktik*, 3. Auflage, Tübingen: Narr.
- ROELCKE, T. (2005): *Fachsprachen*, Berlin: Erich Schmidt.
- SCHRIMPF, U., BAHNEMANN, M. & LECHNER, M. (2019): *Deutsch für Ärztinnen und Ärzte. Trainingsbuch für die Fachspracheprüfung und den klinischen Alltag*, Berlin: Springer.
- SCHRIMPF, U. & BAHNEMANN, M. (2009): *Deutsch für Ärztinnen und Ärzte. Kommunikationstraining für Klinik und Praxis*, Heidelberg: Springer.



- SOPPA, R. (2019): *Ausländische Ärztinnen und Ärzte – Chancen und Herausforderungen im DaZ-Unterricht*, Masterarbeit zur Erlangung des Akademischen Grades, Johannes Gutenberg-Universität Mainz.
- THOMMES, D. & SCHMIDT, A. (2016): *Menschen im Beruf. Medizin*, München: Hueber Verlag.
- VORMWEG, F. (2012): „Sprachstipendienprogramme für Studierende und approbierte Ärzte zur Personalgewinnung – Praxisbeispiel: Münsterlandkliniken“, in: KIRCHNER H., SCHROETER, M. & FLESCHE, M. (Hrsg.): *Personalakquise im Krankenhaus. Ärzte gewinnen, binden, entwickeln*, Berlin/Heidelberg: Springer: 93–94.
- WINTER-HEIDER, C. (2009): *Mutterland Wort: Sprache, Spracherwerb und Identität vor dem Hintergrund der Entwurzelung*, Frankfurt am Main: Brandes & Ape.

## Internetquellen

- ÄRZTEKAMMER NORDRHEIN (2015): *Kommunikation im medizinischen Alltag. Ein Leitfaden für die Praxis*, abrufbar unter: [https://www.aekno.de/fileadmin/user\\_upload/aekno/downloads/leitfaden-kommunikation-2015.pdf](https://www.aekno.de/fileadmin/user_upload/aekno/downloads/leitfaden-kommunikation-2015.pdf), 26.07.2020.
- ÄRZTESTATISTIK 2019, abrufbar unter: <https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/aerztestatistik-2019/>, 25.07.2020.
- ÄRZTESTATISTIK 2016, abrufbar unter: <https://www.bundesaerztekammer.de/presse/pressemitteilungen/news-detail/die-schere-zwischen-behandlungsbedarf-und-behandlungskapazitaeten-oeffnet-sich/>, 25.07.2020.
- BERUFSQUALIFIKATIONSFESTSTELLUNGSGESETZ, abrufbar unter: <https://www.bq-portal.de/Anerkennung-f%C3%BCr-Betriebe/Anerkennungsverfahren/Rechtliche-Grundlagen>, 20.07.2020.
- BÄK (2020a): Bezirksärztekammer Rheinhessen: *Informationen zur Fachsprachprüfung (FSP) für ausländische Ärzte in Rheinland-Pfalz*, abrufbar unter: [https://aerztekammer-mainz.de/pdf/diverse/FSP-Informationen-2020\\_deutsch.pdf](https://aerztekammer-mainz.de/pdf/diverse/FSP-Informationen-2020_deutsch.pdf), 31.07.2020.

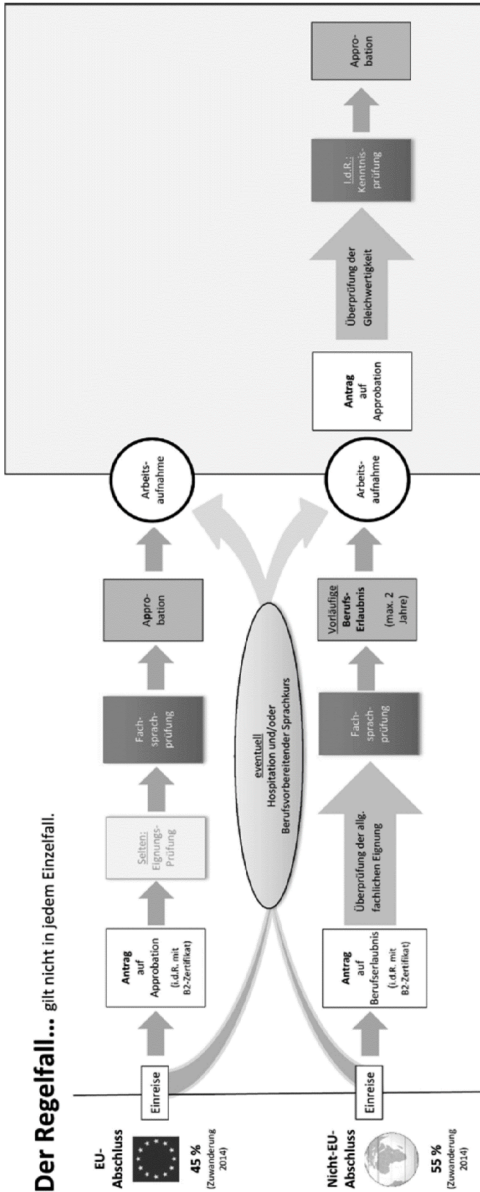
- BÄK (2020b): Bezirksärztekammer Rheinhausen: *Ablaufhinweise zur Fachsprachenprüfung (FSP) in Mainz*, abrufbar unter: <https://aerztekammer-mainz.de/pdf/diverse/FSP-Ablaufhinweise-2020.pdf>, 27.07.2020.
- BfA (2019): Bundesagentur für Arbeit: *Statistik/Arbeitsmarktberichterstattung, Berichte: Blickpunkt Arbeitsmarkt–Fachkräfteengpassanalyse*, Nürnberg, abrufbar unter: [https://statistik.arbeitsagentur.de/Statistikdaten/Detail/201906/arbeitsmarktberichte/fk-engpassanalyse/fk-engpassanalyse-d-0-201906-pdf.pdf?\\_\\_blob=publicationFile](https://statistik.arbeitsagentur.de/Statistikdaten/Detail/201906/arbeitsmarktberichte/fk-engpassanalyse/fk-engpassanalyse-d-0-201906-pdf.pdf?__blob=publicationFile), 31.07.2020.
- BUNDESÄRZTEKAMMER (2017): Deklaration von Genf, abrufbar unter: [https://www.bundesaerztekammer.de/fileadmin/user\\_upload/downloads/pdf-Ordner/International/Deklaration\\_von\\_Genf\\_DE\\_2017.pdf](https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/International/Deklaration_von_Genf_DE_2017.pdf), 14.09.2020.
- BAMF (2015): Bundesamt für Migration und Flüchtlinge: *Fachkräftezuwanderung im internationalen Vergleich, Working Paper 62*, abrufbar unter: [https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/WorkingPapers/wp62-fachkraefte-zuwanderung.pdf?\\_\\_blob=publicationFile&v=11](https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/WorkingPapers/wp62-fachkraefte-zuwanderung.pdf?__blob=publicationFile&v=11), 26.07.2020.
- BMAS (2020): Bundesministerium für Arbeit und Soziales: *Fachkräfteeinwanderungsgesetz*, abrufbar unter: <https://www.bmas.de/DE/Presse/Meldungen/2020/neue-gesetze-fachkraefteeinwanderungsgesetz.html>, 27.07.2020.
- GMK: 87. Gesundheitsministerkonferenz am 26./27. Juni 2014 in Hamburg, TOP 7: *3 Eckpunkte zur Überprüfung der für die Berufsausübung erforderlichen Deutschkenntnisse in den akademischen Heilberufen*, abrufbar unter: [https://www.gmkonline.de/documents/TOP73BerichtP\\_Oeffentl\\_Bereich.pdf](https://www.gmkonline.de/documents/TOP73BerichtP_Oeffentl_Bereich.pdf), 26.07.2020.
- INFORMATIONSPORTAL der Bundesregierung zur Anerkennung ausländischer Berufsqualifikationen, abrufbar unter: <https://www.anererkennung-in-deutschland.de/de/interest/finder/profession>, 20.07.2020.
- MARBURGER BUND (2018): *Deutschkenntnisse – Anforderungen in den Bundesländern für die Approbationerteilung*, abrufbar unter: <https://www.marburger-bund.de/sites/default/files/files/2018-09/deutschkenntnisserman-requirements-approbation.pdf>, 15.09.2020.
- RHEINLAND-INSTITUT, abrufbar unter: <https://unitutor.de/doc/>, 20.07.2020.

STATISTA RESEARCH DEPARTMENT (2019): *Statistiken zum Fachkräftemangel*, abrufbar unter: <https://de.statista.com/themen/887/fachkraeftemangel/>, 21.07.2020.

STATISTA RESEARCH DEPARTMENT (2016): *Prognose zum Fachkräfteengpass in gesundheitswirtschaftlichen Berufen in Deutschland bis zu den Jahren 2020 und 2030*, abrufbar unter: <https://de.statista.com/statistik/daten/studie/703367/umfrage/fachkraefteengpass-in-gesundheitswirtschaftlichen-berufen-in-deutschland-bis-2030/#professional>, 22.10.2020.

TELC: *Gemeinsamer Europäischer Referenzrahmen für Sprachen. Kann-Beschreibungen für Medizinerinnen und Mediziner*, abrufbar unter: [https://www.telc.net/fileadmin/user\\_upload/Broschueren/ger\\_poster\\_medizin.pdf](https://www.telc.net/fileadmin/user_upload/Broschueren/ger_poster_medizin.pdf), 21.07.2020.

Anhang: Anerkennungsverfahren (Deppe et al. 2015: 22f.)





PRÜFEN



## Fachsprachenprüfungen als Nachweis der Sprachkompetenzen?! – Ein Vergleich mit dem neu entwickelten *Sprachtest für ausländische Mediziner (SAM)*

### Einleitung<sup>1</sup>

„Heilkunst braucht auch Sprachkunst“, sagte Dr. Theodor Windhorst, Präsident der Ärztekammer Westfalen-Lippe 2019, anlässlich der 5.000. Fachsprachenprüfung für ausländische Ärzt\*innen (seit ihrer Einführung im Jahr 2014), die von der Kammer abgenommen wurde (vgl. ÄKWL 2019: o.S.). Diese Aussage impliziert zum einen die notwendigen Sprachkompetenzen zur Ausübung eines Heilberufs allgemein; zum anderen stellt sich damit die Frage, welche Sprachkenntnisse dafür benötigt werden bzw. welche Anforderungen dies an eine Fachsprachenprüfung stellt, die diese Kompetenzen bei ausländischen Ärzt\*innen zu überprüfen vermag.

Fest steht, dass „[d]ie Sprache [...] eines der elementaren Werkzeuge für die ärztliche Tätigkeit [ist]“ (Borowski 2018: 17) und „[...] eine Schlüsselfunktion in der Arzt-Patienten-Beziehung“ (Windhorst, zit. in: ÄKWL 2019: o.S.) einnimmt. Fachspezifische Sprachkompetenzen sind nicht nur zur Ausübung des Berufes unabdingbar, sondern in Deutschland auch gesetzlich vorgeschrieben. Gemäß den Beschlüssen der 87. Gesundheitsministerkonferenz 2014 sind allgemeinsprachliche Zertifikate und Diplome zur Überprüfung der fachsprachlichen Deutschkenntnisse nicht ausreichend (vgl. 87.GMK 2014: 2). Fachsprachenprüfungen sollen dafür Abhilfe schaffen und stellen laut Wind-

.....

1 In dem Artikel „Die medizinische Fachsprachenprüfung – Eine Pilotstudie zur Evaluation durch die Teilnehmenden“ (in diesem Band) wurden Desiderata bei der Fachsprachenprüfung thematisiert. Dort wurde der neu entwickelte *Sprachtest für ausländische Mediziner* (kurz SAM) bereits angesprochen, der hier in den Fokus rückt; dabei kommt es zu Doppelungen, die jedoch notwendig sind, um die Artikel auch separat verstehen zu können.



horst einen „Garant für Patientensicherheit, medizinische Versorgungsqualität und Integration“ (zit. in: ÄKWL 2019: o.S.) dar – ein einheitliches Prüfungs- und Anerkennungsverfahren existiert jedoch bislang nicht, trotz Bestrebungen der GMK einer bundeseinheitlichen Vorgehensweise (vgl. 87.GMK 2014: 7).

Der von der LMU bzw. dem Klinikum der Universität München 2018 neu entwickelte *Sprachtest für ausländische Mediziner*<sup>2</sup> (kurz SAM) setzt genau hier an und gilt seitens der Testautor\*innen als „erste[r] und bis dato einzige[r] bundesweite[...] Ansatz einer wissenschaftlich fundierten Fachsprachenprüfung“ (Lenz et al. 2019: o.S.).

Ziel dieses Artikels ist die Vorstellung des SAM und diesen vergleichend zu etablierten Prüfungsformaten zu betrachten, um den Ansatz und die Anforderung einer validen Prüfung der von ausländischen Mediziner\*innen geforderten Fachsprachenkompetenzen zu evaluieren. Dabei wird analysiert, ob der neue SAM den Anforderungen gerecht wird und somit eine Möglichkeit zur bundesweiten FSP darstellen kann.

Dafür wird zunächst ein theoretischer Überblick zu den formalen und inhaltlichen Prüfungsanforderungen sowie deren Herausforderungen gegeben. Dieser dient als Grundlage für die in Kapitel zwei anknüpfenden Analyse und Evaluation der im Bereich Medizin etablierten Prüfungsformate bzw. des SAM. Hierfür werden neben dem SAM exemplarisch die Fachsprachenprüfung Medizin von *telc* und die FSP des Landes Rheinland-Pfalz dargestellt. Abschließend folgt im dritten Kapitel eine Diskussion über die Notwendigkeit und Sinnhaftigkeit, ein neues, einheitliches Prüfungsformat zu etablieren.

## 1 Formale und inhaltliche Anforderungen an Fachsprachenprüfungen

Fachsprachenprüfungen (FSP) lassen sich der Form „fremdevaluative Tests“ (Handt 2002: 187) zuordnen, da ihre Entwicklung, Durchführung und Auswertung durch externe Expert\*innen erfolgt (vgl. ebd.). In der Erwachsenenbildung erfüllen Tests unterschiedliche Funktionen und werden demnach

.....  
2 Im Original ungedeutert.

grundlegend nach Einstufungstest, Diagnostiktest, Lernfortschrittstest und Sprachstandstest unterschieden (vgl. ebd.). Unter Sprachstandstests sind meist standardisierte Sprachzertifikate zu verstehen, die nach einem Sprachbeschreibungsmodell, zunehmend dem GER, den Leistungsstand einer Zweit- oder Fremdsprache zum Zeitpunkt der Prüfung schriftlich und mündlich messen (vgl. ebd.). Der Lernprozess oder -fortschritt wird hierbei nicht festgehalten (vgl. Perlmann-Balme 2016: 420).

Im Bereich Medizin verbessert das Ablegen einer FSP nicht nur die Chancen auf eine Arbeitsstelle, sondern ist mittlerweile gesetzlich verpflichtend, wobei sie dennoch auf eigene Kosten erworben werden muss (vgl. ebd.). Als Ziel der Testsituation wird u.a. beschrieben, „[...] möglichst authentische fremdsprachliche Leistungen der Teilnehmenden herauszufordern“ (Rossa 2016: 400) und zwar unter für alle TN möglichst fairen und vergleichbaren Bedingungen (vgl. ebd.).

Darüber hinaus erheben Sprachtests, wie jede wissenschaftliche Messung, den Anspruch „möglichst objektiv, zuverlässig und gültig“ zu sein (Diekmann 2009: 247). Die drei daraus entstehenden Gütekriterien Objektivität, Reliabilität und Validität spielen gerade bei Sprachstandtests und standardisierten Sprachzertifikaten eine sehr große Rolle (vgl. Handt 2002: 189). Ein valider Test sollte genau das messen, was tatsächlich gefragt ist, z.B. bestimmte Kompetenzen (vgl. ebd.). Ob eine Messung bzw. deren Ergebnisse reliabel, also zuverlässig und genau sind, kann durch ein hohes „Maß an Reproduzierbarkeit“ (Diekmann 2009: 250) gewährleistet werden. Das Erfüllen des dritten Gütekriteriums, die Objektivität, steht eng im Zusammenhang mit Transparenz. Diese wird durch die Offenlegung des methodischen Vorgehens und Darlegung der zu berücksichtigenden Faktoren gewährleistet (vgl. Schmelter 2014: 35f) und stellt die Unabhängigkeit der Testergebnisse von den jeweiligen Bewertenden/Prüfenden dar (vgl. Handt 2002: 189f.). Detaillierte Bewertungskriterien und intensives Prüfer\*innentraining<sup>3</sup> unterstützen die Minimierung des individuellen Bewertungsspielraums (vgl. ebd.).

Ein verstärktes Qualitätsbewusstsein entstand in den 1990er Jahren und zog eine Zunahme an Selbstkontrollen der Sprachtestanbieter mit sich (vgl.

.....

3 Zur Qualität von Prüferqualifizierung siehe Beitrag von Grein in diesem Band.

Perlmann-Balme 2016: 422). Die Vereinigung von Sprachprüfungsanbietern in Europa (*Association of Language Testers in Europe – ALTE*) entwickelte hierfür 17 Mindeststandards<sup>4</sup> sowie im Auftrag des Europarats ein Handbuch zur Entwicklung und Durchführung von Sprachtests unter der Verwendung des *GER* (vgl. ALTE 2012). Der Umsetzung und Einhaltung dieser Qualitätsstandards haben sich ihre Partner\*innen und Mitglieder, vornehmlich international verbreitete Anbieter, verpflichtet. Im Bereich Deutsch als Fremdsprache zählen dazu das Goethe-Institut, die *telc gmbH* und das *TestDaF* Institut. Als Phasen einer standardisierten Prüfung gelten bei *ALTE*: Planung, Testentwicklung, Pilotierung und Informationen von Beteiligten (vgl. ALTE 2012: 79), bspw. auch als Planungsphase, Designphase, Erprobungsphase und Finalisierungsphase (vgl. Shafer & Wiedenkeller 2012: 119ff.) bezeichnet.

Die Bewertung der FSP bzw. die Vergabe dieses Sprachzertifikats ist für die prüfungsteilnehmenden Ärzt\*innen von existentieller Bedeutung, was umso mehr für eine standardisierte und bundesweit einheitliche FSP spricht. Gleichzeitig tragen die Testentwickler\*innen damit eine hohe Verantwortung, die nachzuweisenden Kompetenzen glaubwürdig zu überprüfen und zu dokumentieren, was wiederum an einen hohen Anspruch der Erfüllung der Gütekriterien gekoppelt ist (vgl. Perlmann-Balme 2016: 422). Über diese fachspezifische Sprachkompetenzen gilt es, einen Nachweis zu erbringen, deren Komplexität im folgenden Abschnitt kurz skizziert wird.

## 1.1 Sprachkompetenzen

Medizinische Kommunikation ist durch eine große Vielfalt an Kommunikationsebenen, -formen und -situationen gekennzeichnet, an denen eine Vielzahl an Gesprächspartner\*innen beteiligt sind (vgl. Roelcke 2005: 193)<sup>5</sup>. Zur Bewertung und Überprüfung von Sprachkompetenzen werden meist die Niveaubeschreibungen des *GER* für die Fertigkeiten Hören, Lesen, Schreiben

.....  
4 17 Mindeststandards zur Sicherstellung von Qualität in den Prüfungen der ALTE, abrufbar unter: [https://www.testdaf.de/fileadmin/Redakteur/PDF/TestDaF/ALTE/ALTE\\_Mindeststandards\\_DE.pdf](https://www.testdaf.de/fileadmin/Redakteur/PDF/TestDaF/ALTE/ALTE_Mindeststandards_DE.pdf), 21.07.2020.

5 Für eine ausführlichere Erläuterung der spezifischen Merkmale der medizinischen Fachsprache vgl. Roelcke 2005: 193ff.

und Sprechen herangezogen. Die Stufenbeschreibungen der Globalskala des GER wurden dem medizinischen Kontext für die Niveaustufen B2 und C1 angepasst<sup>6</sup> (Abb. 1, vgl. *telc*<sup>7</sup>). Sie zeigt in Gegenüberstellung die allgemeinen Kann-Beschreibungen auf C1-Niveau und die komplexen Kann-Beschreibungen für die Fertigkeit Sprechen im medizinischen Bereich, was den besonderen Bedarf an spezifischen Sprachkenntnissen verdeutlicht (vgl. Jehle 2019: 49):

	<b>GER:</b> Die Stufen-Beschreibungen der Globalskala	<b>SPRECHEN:</b> Anamnese, Patientenvorstellung, Besprechung
<b>C1</b>	<p>Kann ein breites Spektrum anspruchsvoller, längerer Texte verstehen und auch implizite Bedeutungen erfassen.</p> <p>Kann sich spontan und fließend ausdrücken, ohne öfter deutlich erkennbar nach Worten suchen zu müssen.</p> <p>Kann die Sprache im gesellschaftlichen und beruflichen Leben oder in Ausbildung und Studium wirksam und flexibel gebrauchen.</p> <p>Kann sich klar, strukturiert und ausführlich zu komplexen Sachverhalten äußern und dabei verschiedene Mittel zur Textverknüpfung angemessen verwenden.</p>	<p>Kann sich beinahe mühelos spontan und fließend im klinischen und privaten Alltag ausdrücken. Beherrscht einen großen Wortschatz und kann bei Wortschatzlücken problemlos Umschreibungen finden (Transfersprache Arzt – Patient / Angehörige).</p> <p>Kann Deutsch wirksam und flexibel für soziale und berufliche Zwecke gebrauchen, auch für den Ausdruck von Emotionen (empathisches Sprechen mit Angehörigen, Patienten, z. B. bei Sucht- oder Sozialanamnese), Anspielungen oder zum Scherzen (mit Kollegen in Small Talk-Situationen).</p> <p>Kann komplexe medizinische Sachverhalte oder Verläufe (z. B. pathologische Persönlichkeitsstrukturen oder biografische Besonderheiten) klar und detailliert darstellen.</p> <p>Kann Therapie, OP-Verläufe oder Komplikationen ausführlich beschreiben. Kann untergeordnete Themen integrieren, bestimmte Punkte genauer ausführen und alles mit einem angemessenen Schluss abrunden.</p>

Abb. 1: Auszug aus dem GER, Kann-Beschreibungen für Medizinerinnen und Mediziner (*telc*)

Neben den geforderten Fremd- bzw. Fachsprachenkompetenzen und den vorausgesetzten medizinischen Fachkenntnissen sind für den beruflichen Alltag

.....

6 Das vorgestellte Sprachniveau B2–C1 wurde bewusst gewählt, da es sich dabei um die Kompetenzstufen handelt, die zur Anerkennung im Ausland erworbener fachlichen Qualifikationen im medizinischen Bereich und zum Erhalt der Approbation bzw. Berufserlaubnis in Deutschland als Voraussetzung gelten (vgl. 87. GMK 2014: 3).

7 Für eine zusammenfassende Darstellung vgl. Jehle 2019: 47f. Die vollständige Beschreibung der Kompetenzen von *telc* sind unter folgendem Link zu finden: [https://www.telc.net/file-admin/user\\_upload/Broschueren/ger\\_poster\\_medizin.pdf](https://www.telc.net/file-admin/user_upload/Broschueren/ger_poster_medizin.pdf), 21.07.2020.

und letztlich dessen Erfolg weitere Kompetenzen essenziell. Im Tätigkeitsfeld von Ärzt\*innen ist der multikulturelle Kontakt und Austausch mit Menschen allgegenwärtig, sodass ausgeprägte Sozialkompetenzen<sup>8</sup> und interkulturelle Kompetenzen sowie berufsübergreifende Schlüsselqualifikationen bzw. sog. *soft skills* benötigt werden. Die Beherrschung dieser Fähigkeiten manifestiert sich immer auch sprachlich, folglich lässt sich eine ‚sprachliche Handlungskompetenz‘ als übergeordnetes Ziel definieren. Handlungskompetenz wird dabei „[...] als die Bereitschaft und Befähigung des Einzelnen [verstanden], sich in beruflichen, gesellschaftlichen und privaten Situationen sachgerecht, durchdacht sowie individuell und sozial verantwortlich zu verhalten“ (KMK 2018: 15). Zu diesen sprachlichen Fähigkeiten, sich als Ärztin/Arzt je nach Kontext und Gesprächspartner\*in angemessen und seine Sprechabsichten adäquat auszudrücken, gehören auch nonverbale und paraverbale Aspekte der Kommunikation<sup>9</sup>, denen in verschiedenen Kommunikationsmodellen<sup>10</sup> Rechnung getragen wird. Diese können wiederum ebenfalls in unterschiedlichen kulturellen Räumen divergieren (vgl. Kellnhauser & Schewior-Popp 1999: 23ff). All diese Aspekte der ‚Sprachkompetenz‘ sollten m.E. auf dem Weg zur Berufsankennung bzw. -ausübung eine Rolle spielen. Hierbei stellt sich die Frage, inwieweit diese im Rahmen einer FSP bewertbar sind. Die Komplexität der medizinischen (Fach-)Sprache stellt die Prüfungsentwicklung zur Überprüfung dieser Aspekte vor große Herausforderungen. In den Beschlüssen der 87. GMK sind die Mindestanforderungen der FSP verankert und stellen die inhaltliche Grundlage dieser Prüfung(en) dar.

.....

- 8 Zur Diskussion der Berücksichtigung von Sozialkompetenzen als Auswahlkriterium bei Medizinstudierenden siehe bspw. <https://blogs.faz.net/blogseminar/wie-misst-man-soziale-kompetenz/>, 30.09.2020.
- 9 Unter nonverbaler Kommunikation werden verstanden: Blick, Blickkontakt, Blickwechsel; Körperkontakt; Körperhaltung, Kleidung; räumliche Distanz; Mimik, Gestik. Als paralinguistische Kommunikation bezeichnet man die Art und Weise, wie eine Botschaft weitergegeben wird. Dazu gehören: Lautstärke, Sprechgeschwindigkeit, Sprechpausen, Stimmvariationen“ (Kellnhauser & Schewior-Popp 1999: 23).
- 10 Für weitere Informationen vgl. bspw.: Emmerling, P. (2015): *Ärztliche Kommunikation. Als Erstes heile mit dem Wort...*, Stuttgart: Schattauer; Baller, G. & Schaller, B. (2017): *Kommunikation im Krankenhaus. Erfolgreich kommunizieren mit Patienten, Arztkollegen und Klinikpersonal*, Berlin/Heidelberg: Springer; Bechmann, S. (2014): *Medizinische Kommunikation. Grundlagen der ärztlichen Gesprächsführung*, Tübingen: Narr.

## 1.2 Richtlinien des GMK-Beschlusses

Die Beschlüsse der GMK gelten nicht nur als gesetzliche Grundlage der zur Berufsausübung erforderlichen Sprachnachweise, sondern geben auch Aufschluss über den inhaltlichen und zeitlichen Rahmen des abzulegenden Sprachtests. Diese gelten für Ärzt\*innen, Zahnärzt\*innen, Apotheker\*innen sowie psychologische Psychotherapeut\*innen und Kinder- und Jugendlichenpsychotherapeut\*innen. Welche Kompetenzen die jeweilige Berufsgruppe mündlich wie schriftlich beherrschen muss und letztlich in der FSP überprüft werden sollen, sind ähnlich wie die in Kapitel 1.1 beschriebenen Kann-Beschreibungen des GER formuliert. Zusammengefasst ist dort für die Gruppe der Ärzt\*innen und Zahnärzt\*innen Folgendes zu finden (vgl. 87.GMK 2014: 3):

Personengruppen bzw. Gesprächspartner\*innen: Patient\*innen, Angehörige, Kolleg\*innen und Angehörige anderer Berufe

Arbeitssituationen: Anamnese; Information und Aufklärung über Befunde, Erkrankungen, Behandlung(sverlauf) und geplante Maßnahmen; Patient\*innenvorstellung; Krankenunterlagen führen

Sprachhandlungen: Gesprächspartner\*innen verstehen, sich spontan und fließend verständigen, informieren, aufklären, beschreiben, darstellen, sich klar und detailliert ausdrücken, ausstellen

Dabei wird die Patienten\*innensicherheit hervorgehoben, indem Missverständnisse, Fehldiagnosen, falsche Therapieentscheidungen und Therapiefehler ausgeschlossen werden sollen.

Daraus resultieren folgende Mindestanforderungen für den Sprachtest (87. GMK 2014: 6):

Der Sprachtest umfasst

- ein simuliertes Berufsangehöriger-Patienten-Gespräch, in dem die unter I.1 in Bezug auf die Kommunikation zwischen Berufsangehörigen und Patienten beschriebenen Anforderungen unter Beweis gestellt werden (20 Minuten),

- das Anfertigen eines in der ärztlichen, zahnärztlichen, pharmazeutischen oder psychotherapeutischen Berufsausübung üblicherweise vorkommenden Schriftstückes (z.B. Kurz-Arztbrief) zum Nachweis der unter I.1 beschriebenen schriftlichen Sprachanforderungen (20 Minuten),
- ein Gespräch mit einem Angehörigen derselben Berufsgruppe, [...] zum Nachweis der unter I.1. beschriebenen Anforderungen in Bezug auf die Zusammenarbeit mit Kolleginnen und Kollegen oder im Team (20 Minuten).

Hier steht v.a. die Überprüfung des Hörverstehens sowie die mündliche und schriftliche Ausdrucksfähigkeit im Vordergrund. Es wird betont, dass das (medizinische) Fachwissen der Prüfungsteilnehmenden nicht überprüft werden darf.

Außerdem wird festgelegt, dass der Sprachtest in Form einer Einzelprüfung stattfinden und die Bewertung durch mindestens zwei Prüfer\*innen erfolgen muss. Davon hat mindestens eine Person Angehörige\*r der Berufsgruppe zu sein, die auch der oder die zu Prüfende angehört und soll Deutsch als Muttersprache beherrschen (vgl. ebd.). Über das Bestehen des Sprachtests entscheidet das Bewertungsgremium. Welche Bewertungskriterien oder -formen dabei gelten, ist im Beschluss nicht weiter ausgeführt. Im Falle des Nicht-Bestehens ist ein beliebiges Wiederholen des ganzen Tests möglich (vgl. ebd.: 7).

Grundsätzlich werden die Sprachtests bei der jeweiligen Ärztekammer oder Approbationsbehörde des Bundeslandes abgelegt, aber es wird auch die Möglichkeit eingeräumt, andere Nachweise anzuerkennen, solange die beschriebenen Anforderungen damit belegt werden können. Liegt eine amtliche Bescheinigung der erforderlichen (Fach)Sprachkenntnisse vor, sind die anderen Länder verpflichtet, diese auch zu akzeptieren. Insgesamt drückt die GMK aber das Bestreben einer bundeseinheitlichen Vorgehensweise aus (vgl. ebd.).

Mit dem Inkrafttreten dieses Beschlusses entstand der Bedarf an Überprüfungsmöglichkeiten der geforderten medizinischen Fachsprache und somit an einer Fachsprachenprüfung (FSP). Der Entwicklung von Deutsch-Prüfungen im Bereich Medizin nahm sich als Institution die *telc gmbH* an, wobei die Konzeption und Durchführung in erster Linie an die jeweiligen Landesärzte-

kammern in ‚Eigenregie‘ übertragen wurde. Angesichts dessen kann hier nicht von einer bundeseinheitlichen Prüfungsvorgehensweise gesprochen werden. Zu diesem Schluss kommen auch Malakhova & Lenz nach Überprüfung dieser Hypothese anhand von Daten über Ablauf, Dauer, Gebühren, Teilnehmenden und Durchfallquoten für die Jahre 2014–2017, trotz teilweise mangelhafter Zugänglichkeit der Daten seitens der Ärztekammern (vgl. 2018: 1) und fordern ebenfalls eine Vereinheitlichung der FSP (vgl. ebd.: 3).

Die durch die GMK reglementierte FSP hat zum Ziel, zugewanderte Ärzt\*innen einheitlich zu prüfen, um eine kompetente Sprachverwendung im ärztlichen Berufsfeld festzustellen. Jedoch sind Prüfungen dieser Art nicht unumstritten und bergen Herausforderungen und Schwierigkeiten, worauf sich der folgende Abschnitt konzentriert und den Blick auf die Prüfungspraxis einbezieht.

### 1.3 Herausforderungen und Probleme der FSP

Die Herausforderungen bei der Entwicklung einer validen Fachsprachenprüfung bestehen aus der Berücksichtigung von Elementen aus dem jeweiligen Fachbereich, der Testmethodik, der Fach- und Fremdsprachendidaktik (Medizin und Deutsch als Fremdsprache) und der Psychometrie (vgl. Lenz et al. 2019: o.S.).

Hinsichtlich des in Kapitel 1 beschriebenen Ziels einer authentischen Prüfung entsteht ein „methodisches Dilemma“ (Rossa 2016: 400):

Je komplexer die erfasste Leistung ist, die in subjektiv bedeutsamen Kontexten und ggf. in Interaktion mit mehreren [Sprechenden, N. J.] erfasst wird, desto eher kann sie als repräsentatives Beispiel für die Kompetenzen der Testteilnehmenden gelten. Mit der Komplexität der Leistung steigt aber auch die Schwierigkeit, die Messung reliabel und objektiv durchführen zu können (ebd.),

da sich die individuellen Äußerungen der Prüfungsteilnehmenden u.U. nicht mehr eindeutig den Kategorien eines Bewertungsschemas zuordnen lassen (vgl. ebd.). Folglich schaffen detaillierte Bewertungskriterien Abhilfe. Je spe-



zifischer diese sind, „desto schwieriger ist es jedoch [...], alle Kategorien bei der Testdurchführung gleichermaßen im Blick zu behalten“ (Handt 2012: 190). Besonders bei der Bewertung mündlicher Leitungen sei es kaum möglich, als Bewerter\*in „auf mehr als drei oder vier Kriterien gleichzeitig zu achten“ (ebd.) wie bspw. „Berücksichtigung (vorgegebener) Leitpunkte, kommunikative Gestaltung und formale Richtigkeit“ (ebd.). Bei der FSP übernehmen die Prüfenden in der Regel auch die Rolle der/des Gesprächspartnerin/-partners eines simulierten Prüfungsgesprächs, was die Konzentration auf die Bewertung erschwert. Dabei nimmt gerade die reliable Bewertung der produktiven Leistungen hinsichtlich der Testgüte eine Schlüsselfunktion ein, was nur durch geschulte Bewerter\*innen sichergestellt werden kann (vgl. Perlmann-Balme 2016: 423). Dies verlangt neben einer standardisierten Prüfer\*innenschulung auch eine regelmäßige und systematische Kontrolle seitens der Testanbieter (vgl. ebd.)

Ein weiterer Punkt ist die mangelnde Transparenz, da es seitens der Ärztekammern kaum detaillierte Veröffentlichungen der Bewertungskriterien der vorhandenen Fachsprachenprüfungen gibt, was zu einer schlechteren Vorbereitungsmöglichkeit seitens der Teilnehmenden, aber auch der Prüfenden führt. Demnach hängt die Bewertung stark von deren Einschätzungen der Kenntnisse der/des Prüfungsteilnehmenden ab (vgl. Dippold-Schenk 2018: 41). Kritisiert wird außerdem, dass die FSP fast ausschließlich von Ärzt\*innen konzipiert wird und keine Zusammenarbeit von Sprachwissenschaftler\*innen und Mediziner\*innen stattfindet (vgl. Deppe & Borowski et al. 2015: 27). Dies betrifft auch die Zusammensetzung der Prüfungskommission der Ärztekammern: Ärzt\*innen sind obligatorisch, Sprachwissenschaftler\*innen nicht (vgl. 87. GMK 2014: 6). Dabei wird angenommen, dass es aufgrund dessen zu einer Vermischung von fachsprachlichen Merkmalen und Sprachkenntnissen kommt (vgl. Deppe & Borowski et al. 2015: 37). Die alleinige Bewertung durch Ärzt\*innen könne dagegen zu einer weniger fundierten Beurteilung führen, da Mediziner\*innen nicht über die „angemessenen Diagnoseinstrumente“ (ebd.) verfügen. Während der Prüfung würden so eher die medizinischen Aussagen bewertet und es werde vom sprachlichen Schwerpunkt abgekommen (vgl. ebd.).

Bei dieser Art eines offenen Prüfungsformats, bei dem die Sprachproduktion im Rahmen der Aufgaben ‚frei‘ erfolgt, wird die „face validity“ (Handt 2012: 190) angeführt, was bedeutet, dass „eine hohe Entsprechung zwischen

dem Testformat und der Art von Verstehen oder Kommunikation, die man abtesten will, [besteht, N. J.]“ (ebd.). Diese fordert auch sog. „integrierte Aufgabenformate“ (ebd.), was eine Verbindung der Teilfertigkeiten, z.B. Lesen und Sprechen, zu Gunsten einer authentischen Kommunikation bedeutet. Bei einer solchen Testkonstruktion lässt sich jedoch die Ursache für weniger gute Leistungen schwieriger nachvollziehen, da sie auf Probleme beim Verstehen oder der aktiven Produktion zurückgeführt werden können (vgl. ebd.).

Kritik an den Prüfungen, die im Zuge der GMK-Beschlüsse anfänglich zur Verfügung standen, kam auch von Roche. Einerseits nehmen die Prüfungen

die medizinische Umgangssprache und berufssprachliche Kompetenzen selektiv in den Blick, umfassen in der Regel nur begrenztes Fachvokabular, behandeln wenige schriftliche Textsorten und so gut wie keine wissenschaftliche, medizinische Fachsprache (Roche 2014: 317),

andererseits sind die Sprachtests meist geringer auf Fach- und Berufssprache ausgelegt und daher in Bezug auf medizinsprachliche Kommunikationskompetenz nur wenig aussagekräftig (vgl. ebd.). Des Weiteren merkt er an, dass „viele Tests und Materialien immer noch ad hoc und ohne wissenschaftlich belastbare Systematik entwickelt [werden, N. J.]“ (ebd.).

Neben der Entwicklung betreffen die meisten der aufgeführten Schwierigkeiten und Herausforderungen der FSP die praktische Durchführung, nicht zuletzt aufgrund ihrer unterschiedlichen Ausführungen.

Eine Auflistung des Marburger Bunds (2018) gibt eine Übersicht der im jeweiligen Bundesland anerkannten bzw. zu absolvierenden FSP. Diese verdeutlicht, dass von einer „undurchsichtigen und schwer durchschaubaren“ (ebd.) Prüfungslandschaft gesprochen werden kann. Darauf zurückzuführen ist der sich entwickelnde bzw. bereits existierende ‚Prüfungstourismus‘ (vgl. Malakhova & Lenz 2018: 1). Der Prüfungsort ist vom Wohnort bzw. dem offiziellen Wohnsitz der Teilnehmenden abhängig. In Internetforen<sup>11</sup>, einschlägigen Facebook- oder Whatsapp-Gruppen und auch innerhalb von Prüfungsvorbereitungskursen herrscht ein reger Austausch über praktische Erfahrungen

.....

11 Vgl. bspw. <http://approbations.info/blog/>, 28.07.2020.

und Empfehlungen sowie darüber, wo die FSP vermeintlich am einfachsten ist. Ausländische Ärzt\*innen ziehen sogar in ein anderes Bundesland, um die Chancen auf das Bestehen der Prüfung und damit einer beruflichen Perspektive in Deutschland zu erhöhen<sup>12</sup>. Ist einmal eine FSP bestanden, sind alle Bundesländer laut den GMK-Beschlüssen dazu verpflichtet, diese auch anzuerkennen (vgl. 87. GMK 2014: 7). Aufgrund der damit einhergehenden langwierigen Bürokratie bestimmt jedoch oft der Ort der Prüfung auch den Arbeitsort, womit dem Ärzte\*mangel in bestimmten Regionen sicherlich nicht geholfen ist. Ein einheitliches Prüfungssystem würde dem Prüfungstourismus entschieden entgegenwirken.

Die Probleme und Herausforderungen lassen sich wie folgt zusammenfassen:

1. Konzeption und Entwicklung: Berücksichtigung aller beteiligten Fachbereiche
2. Test und Material: wissenschaftlich fundiert
3. offenes Prüfungsformat: Verbindung der Teilfertigkeiten  
→ Ursache für weniger gute Leistungen schwieriger nachvollziehbar
4. Ziel: authentische fremd-/fachsprachliche Leistung führt zu Komplexität der Leistung
5. Rollenübernahme der Gesprächspartner\*innen bei Prüfungsteilen seitens der Prüfenden  
→ Schwierigkeit einer reliablen und objektiven Messung
6. geschulte Bewerter\*innen = Schlüsselfunktion für reliable Bewertung der produktiven Leistungen  
→ standardisierte Prüfer\*innenschulung/Kontrolle seitens Testanbieter
7. Prüfungskommission: Zusammensetzung meist nur aus Ärzt\*innen, keine Sprachwissenschaftler\*innen (DaFler\*innen)  
→ weniger fundierte Beurteilung/Bewertung mit sprachlichem Schwerpunkt

.....

12 Diese Informationen stammen aus informellen Gesprächen (im Juli 2020) mit einem Kurs-träger in Mainz, deren Kursteilnehmenden darüber berichteten und dies als Grund des Kursabbruches angaben.

8. Transparenz: Zugang zu Material zur Vorbereitung (z.B. Modelltests) und Bewertungskriterien etc.
9. uneinheitliches Prüfungssystem: unterschiedliche Ausführungen je nach Bundesland  
→ Prüfungstourismus
10. Anerkennungspflicht aller FSP seitens der Bundesländer, aber langwierige Bürokratie  
→ wirkt Deckung des Ärzte\*mangels in bestimmten Regionen entgegen

Die Testentwickler\*innen des SAM erklären als Ziel, „den Sprachtest mittelfristig als Beitrag zu einer nachhaltigen Verbesserung gegenwärtiger Testverfahren zum Einsatz zu bringen, nicht nur in Bayern“ (Lenz et al. 2019: o.S.). Was der SAM im Vergleich zu den etablierten FSP leistet, gilt es demnach näher zu betrachten.

Im Folgenden wird zunächst das Forschungsdesign präsentiert (Kapitel 2), auf dessen Grundlage meine Analyse durchgeführt wurde. Daran anschließend werden die exemplarisch ausgewählten FSP sowie der SAM in ihren Grundzügen vorgestellt (Kapitel 2.1) und im darauffolgenden Kapitel (2.2) evaluiert.

## 2 Analysevorhaben und Forschungsdesign

Das Vorhaben, einen exemplarischen Vergleich vorhandener Prüfungsformate mit dem neu entwickelten SAM durchzuführen, erfolgt mittels eines Daten- und Methoden-Mix. Als Methoden wurden hierbei die Dokumentenanalyse und ein (Experten-)Interview eingesetzt. Für eine Bestandsaufnahme sowie die Analyse der Prüfungen dienten fachrelevante Dokumente v.a. seitens der Prüfungsanbieter bzw. Ärztekammern und einer der SAM-Testentwickler\*innen als Daten- bzw. Informationsquellen.

Die Dokumentenanalyse sollte Aufschluss zu den Prüfungsdetails geben. Aus den Erkenntnissen der Inhaltsanalysen wurde für das Interview ein Leit-

faden erstellt. Das mit einem der Testentwickler des SAM geführte Interview<sup>13</sup> diente zur Vertiefung und ermöglichte, die Perspektive der Testentwicklung einzubeziehen.

Der Vergleich und die Analyse der Prüfungen erfolgte auf der Basis der in Kapitel 1 vorgestellten Aspekte hinsichtlich der testtheoretischen und inhaltlichen Anforderungen an die FSP. Hierfür werden die exemplarisch angeführten Prüfungen in Bezug auf folgende Punkte hin analysiert: Aufbau, Anforderungen, Ablauf, Material, Bewertung, Prüfer\*innen. Die Bewertungsgrundlage stellen dabei die Erfüllung der Gütekriterien und die inhaltliche Umsetzung der GMK-Beschlüsse dar, die letztlich über die Testgüte entscheiden. Im Anschluss sollte damit auszugsweise geklärt werden können, ob die etablierten FSP bzw. der neu entwickelte SAM den Test-Anforderungen gerecht werden und sie als valider Nachweis der komplexen Sprachkompetenzen gelten können. Zunächst werden dafür die drei Testformate in ihren Grundzügen vorgestellt, bevor diese nach den o.a. Kriterien analysiert werden.

## 2.1 Prüfungsvorstellung

In den folgenden Abschnitten werden exemplarisch zwei etablierte und anerkannte Fachsprachenprüfungen, die *telc Deutsch B2·C1 Medizin Fachsprachprüfung* und die rheinland-pfälzische FSP sowie der neu entwickelte *Sprachtest für ausländische Mediziner (SAM)* präsentiert.

### 2.1.1 *telc Deutsch B2·C1 Medizin Fachsprachprüfung*

*telc* ist eine gemeinnützige gGmbH sowie ein Tochterunternehmen des *Deutschen Volkshochschul-Verbands e. V.* mit Zentrale in Frankfurt a. M. und zählt zu den führenden Anbietern standardisierter Sprachprüfungen, die international anerkannt sind. *telc* bietet über 90 zertifizierte Sprachtests in zehn Sprachen auf allen Kompetenzstufen des *GER* an. Die Testauswertung und Zertifizierung

.....

13 Das Interview wurde mit Holger Lenz im Juli 2020 geführt. Er ist wissenschaftlicher Mitarbeiter am Institut für Didaktik und Ausbildungsforschung in der Medizin am Universitätsklinikum der LMU in München. Dort leitet er u.a. das OFIF-Projekt zur Förderung der Kommunikationsfähigkeit internationaler Studierender der Humanmedizin und war an der Entwicklung des SAM beteiligt.

dieser Prüfungen als auch die Schulung und Lizenzierung der Prüfer- und Bewerter\*innen gehören zu ihren Aufgaben (vgl. *telc-a*). Im Bereich Heilberufe wurden vier verschiedene Prüfungen entwickelt: drei im Bereich Medizin und eine im Bereich Pflege auf B1/B2-Niveau (vgl. *telc-b*). Im Rahmen dieses Beitrags wird lediglich die zur Beantragung der Approbation bzw. Berufserlaubnis<sup>14</sup> anerkannte *telc Deutsch B2-C1 Medizin Fachsprachprüfung* betrachtet.

Diese orientiert sich laut *telc* eng am Beschluss der 87. GMK (vgl. *telc-c*). Folglich werden sowohl allgemeinsprachliche Kenntnisse auf B2 als auch fachsprachliche auf C1-Niveau geprüft. Die Kenntnisse werden in allen vier Fertigkeiten, Hören, Lesen, Schreiben und Sprechen abgefragt (ausführlich im Anhang). Der schriftliche Teil umfasst die drei Aufgaben Hörverstehen, Lesen und Sprachbausteine und ist bei maximaler Bearbeitungszeit von 80 Minuten im Gegensatz zu den anderen *telc*-Prüfungen im Bereich Medizin nicht so umfangreich. Bei der mündlichen Prüfung handelt es sich um eine 60-minütige Einzelprüfung, in der die Schreibkompetenz (das Verfassen eines Kurz-Arztbriefs) unter einer Vorbereitungszeit von fünf Minuten integriert ist. Beide Prüfungsteile sind dabei gleichgewichtet, wobei bei der Einzelprüfung die mündlichen Teile 1 und 3 zusammen mit maximal 35 Punkten bewertet werden.

Zum Bestehen dieser Prüfung sind im schriftlichen und mündlichen Teil je 60%, sprich 30 Punkte von maximal 50 erreichbaren je Teil, notwendig (vgl. Fachsprachprüfung Übungstest 1).

Für alle Prüfungen gelten die allgemeinen Bewertungskriterien von *telc*, die sowohl sprachliche als auch inhaltliche Angemessenheit berücksichtigen: Für den schriftlichen Teil (Kurz-Arztbrief) werden *inhaltliche Angemessenheit/Aufgabenbewältigung, Kommunikative Gestaltung, Korrektheit* und *Wortschatz* bewertet (vgl. ebd.: 37f.). Beim mündlichen Teil sind das: die *inhaltliche Aufgabenbewältigung* und in Bezug auf die Sprache: *Aussprache/Intonation, Flüssigkeit, Korrektheit* und *Wortschatz* (vgl. ebd.: 39f.). Die Bewertung des schriftlichen Teils der Prüfung kann nur eine approbierte Ärztin/ein approbierter

.....

14 Mit einer Berufserlaubnis kann im Krankenhaus oder einer Praxis gegen Bezahlung gearbeitet werden (max. zwei Jahre). Eine Weiterbildung zu einer Facharztqualifikation ist erst nach Erhalt der Approbation möglich (vgl. Bezirksärztekammer 2020a: 2).

Arzt oder die/der Inhaber\*in einer *telc* Bewerterlizenz<sup>15</sup> übernehmen. Äquivalent dazu wird die mündliche Prüfung von approbierten Ärzt\*innen oder *telc*-lizenzierten Prüfer\*innen abgenommen, wobei die/der approbierte Ärztin/Arzt sowohl die Rolle der Patientin/des Patienten bei der Anamnese als auch der/des Kolleg\*in bei der Fallvorstellung übernimmt. Die jeweilige Gesamtprüfung kann beliebig oft wiederholt werden, wobei die Kosten für die Prüfung vom jeweiligen Testzentrum selbst festgelegt werden (vgl. *telc* Deutsch-Prüfungen für Ärztinnen und Ärzte)<sup>16</sup>. Diese Prüfung ist als offizieller Nachweis in Hessen, Saarland und Schleswig-Holstein anerkannt (vgl. Marburger Bund 2018).

### 2.1.2 Ärztliche Fachsprachenprüfung Rheinland-Pfalz<sup>17</sup>

Die ärztliche Fachsprachenprüfung in Rheinland-Pfalz wird im Auftrag der Landesärztekammer Rheinland-Pfalz und des Landesamtes für Soziales, Jugend und Versorgung (LSJV) in Koblenz von der Bezirksärztekammer Rheinhessen durchgeführt, von der aktuell 4.800 Ärzt\*innen betreut werden (vgl. BÄK: o.J.). Die dort durchgeführte Prüfung ist die einzige in RLP zur Beantragung der Approbation bzw. Berufserlaubnis anerkannte FSP (vgl. Marburger-Bund 2018: 4). Der Wohnort bzw. das Nachweisen eines konkreten Arbeitsangebots ist Bedingung für das Ablegen der Prüfung in RLP bzw. dem jeweiligen Bundesland. Die Anmeldung zur Prüfung ist nur nach Erhalt einer qualifizierten Eingangsbestätigung des LSJV im Zuge der Beantragung auf Approbation bzw. Berufserlaubnis möglich. Zudem sollte ein Sprachzertifikat vorliegen<sup>18</sup>, wobei das Niveau (nach *GER*) hierbei nicht spezifiziert ist.

.....  
15 Bei *telc* wird zwischen einer Bewerterlizenz für das Bewerten schriftlicher Leistungen und einer Prüferlizenz für die mündlichen Prüfungsteile unterschieden. Weitere Informationen siehe unter: <https://www.telc.net/lehrkraefte/pruefer-und-bewerter/telc-lizenz-erwerben.html>, 25.07.2020.

16 Mehr Informationen zu allen *telc*-Prüfungen im Bereich Heilberufe sind abrufbar unter: <https://www.telc.net/medizin-pflege/kurstraeger-lizenzpartner/sprachpruefungen/medizin.html>, 20.09.2020.

17 Vgl. die Studie von Jehle zur FSP Rheinland-Pfalz in diesem Band.

18 Hier liegt eine deutliche Diskrepanz zwischen der Auflistung der für den FSP-Antrag benötigten Unterlagen auf der Informationsseite der Ärztekammer und der praktischen Handhabung vor. Siehe dazu Umfrage im Beitrag von Jehle in diesem Band.

Nach erfolgreicher Anmeldung erwartet die Prüfungsteilnehmenden eine 2,5–3 stündige Absolvierung (inkl. Wartezeiten) mit fünf bzw. sechs Teilen und mehreren Abschnitten, die unterschiedliche Wertigkeit für die Gesamtprüfung haben<sup>19</sup>:

1. Verständnistest (vorher Vokabeltest): Hier werden je 20 medizinische Begriffe aus der Fachsprache in das Deutsche von Patient\*innen übersetzt und umgekehrt sowie eine Schemazeichnung des Körpers mit medizinischen Bezeichnungen beschriftet bzw. ausgewählte Körperteile (in Laiensprache) unterschiedlich markiert. Für die Vorbereitung dieser Aufgaben wird die Arbeit mit medizinischen Wörterbüchern empfohlen.
2. Freies Gespräch mit folgender Erläuterung: „Hier werden die Prüfer Ihnen Fragen stellen, um zu beurteilen, wie gut Sie Sprache verstehen und wie gut Sie Antworten formulieren können“.
3. Ein 20-minütiges Arzt\*-Patienten\*-Gespräch (Anamnesegespräch/Patienten\*aufklärung) in laienverständlicher Alltagssprache. Nicht eine vollständige Anamnese, sondern das Sprechen, Verstehen und Reagieren auf Patient\*innen steht hierbei im Vordergrund und führt bei Nichtbeachten der Aussagen und Fragen dieser zu erheblichem Punkteverlust.
4. Abfassen eines Schriftstücks (Kurz-Arztbrief): Hier soll eine aktuelle Anamnese in Form eines Fließtexts geschrieben und die erfragten Informationen aus dem vorangegangenen Anamnesegespräch stichwortartig erfasst werden. In Fachsprache folgen Ausführungen zu Verdachts- und Differenzialdiagnose(n) und zur zuvor mit dem Patient\*in mündlich besprochenen Untersuchungsforderung. Hierfür bekommen die Prüfungsteilnehmenden einen vierseitigen Anamnesebogen zum Ausfüllen, der auf der Internetseite zur

.....

19 Die Ausführungen beruhen auf den auf der Internetseite der Bezirksärztekammer veröffentlichten Angaben (vgl. BÄK 2020a). Trotz zusätzlicher „Ablaufhinweise zur FSP in Mainz“ (vgl. BÄK 2020b) geben sie nicht umfassend Aufschluss z.B. zu Details der Prüfungsteile, Bewertungskriterien und Prüfer\*innen. Auf meine Anfrage bei der Kammer wurde leider nicht reagiert.



Vorbereitung einsehbar und herunterladbar ist<sup>20</sup>. Inkludiert ist das Ausfüllen von u.U. davon unabhängigen Anforderungsscheinen für weiterführende Untersuchungen.

5. Aufgaben zum Lese- und Hörverstehen sowie ein Arzt\*-Arzt\*-Gespräch (was nach Hinzufügen des ‚freien Gesprächs‘ eigentlich Teil 6 darstellen müsste<sup>21</sup>): Hier werden zu einem einseitigen Text (z.B. Befunde oder Arztbrief) drei Fragen auf einem Vordruck schriftlich beantwortet. Außerdem werden per Telefon (zwei Anrufe) patientenbezogene Informationen (z.B. aus dem Labor) übermittelt, die im Prüfungsbogen zu notieren sind. Bei dem Gespräch von Ärztin/Arzt zu Ärztin/Arzt handelt es sich um ein Fachgespräch, bei dem in berufstypischer Sprache ein Bericht über die/den Patient\*in aus Teil 1 oder die schriftlich bzw. akustisch übermittelten Informationen aus Teil 4 oder sonstige medizinische Themen erfolgt. An dieser Stelle wird betont, dass es nicht um medizinisches Fachwissen und dessen Überprüfung geht, sondern um „die Fähigkeit, sich gut auszudrücken und sich fachsprachlich miteinander verständigen zu können“. Medizinische Fehler oder fehlendes medizinisches Wissen werden nicht bewertet.

Die vorgestellten Prüfungsteile können hierbei in unterschiedlicher Abfolge geprüft werden (vgl. BÄK 2020a: 2). Dies übernehmen in der Regel zwei approbierte Ärzt\*innen, die auch die Rolle der Gesprächspartner\*in (Patient\*in und Kolleg\*in) einnehmen. Für das Bestehen muss jeder Prüfungsteil mit einer ihm zugeordneten Mindestpunktzahl abgeschlossen werden, wobei das Nichterreichen der erforderlichen Mindestpunkte in einem Teil zum sofortigen Abbruch der Prüfung führt. Die Prüfung gilt insgesamt als bestanden, wenn die Gesamtpunktzahl aus allen Prüfungsteilen bei über 60% liegt, d.h. eine Addition der

.....

20 Anamnesebogen verfügbar unter: <https://aerztekammer-mainz.de/pdf/diverse/FSP-Anamnesebogen-2020.pdf>, 11.08.2020.

21 Letzte Änderungen des Prüfungsablaufs bzw. der -teile erfolgten im Juni/Juli 2020. Zuvor: 1. Vokabeltest (ohne Körperschema), 2. Arzt\*-Patienten\*-Gespräch, 3. Verfassen eines Schriftstücks (Kurz-Arztbrief), 4. Lese- und Hör-Verstehen, 5. Arzt\*-Arzt\*-Gespräch. Gründe für die Ergänzungen sind nicht genannt.

Mindestpunktzahlen aus allen Prüfungsteilen ist nicht ausreichend und muss durch höhere Punktzahlen in anderen Teilen ausgeglichen werden (vgl. ebd.). Das Prüfungsergebnis wird direkt im Anschluss an die Prüfung (mündlich) bekannt gegeben und bei bestandener Prüfung eine Bescheinigung (im Original) ausgestellt. Im Falle des Nichtbestehens kann die Prüfung wiederholt werden. Die Kosten für die Prüfung liegen derzeit bei 425€ (vgl. ebd.: 1).

### 2.1.3 SAM – Sprachtest für ausländische Mediziner

Beim *Sprachtest für ausländische Mediziner (SAM)* handelt es sich um eine FSP, die an der LMU München in Kooperation mit dem Klinikum der Universität von einem Expert\*innen-Team mehrerer Fachdisziplinen (Didaktik und Ausbildungsforschung in der Medizin, Deutsch als Fremdsprache, Empirische Pädagogik und pädagogische Psychologie) mit der Unterstützung und Förderung des Bayerischen Staatsministerium für Gesundheit und Pflege (StMGP) 2018 entwickelt und bereits pilotiert wurde. Das Ziel war dabei die „Entwicklung einer wissenschaftlich fundierten und standardisierten Fachsprachenprüfung für das Bundesland Bayern“ (Lenz et al. 2019: o.S.), die „in situativen Prüfungsstationen [...] fachsprachliche und kommunikative Kompetenzen auf C1-Niveau abprüfen [soll, N. J.]“ (ebd.). Voraussetzung zur Ablegung der Prüfung ist ein allgemeinsprachliches B2-Zertifikat. Die Vorgaben der GMK dienen auch bei dieser FSP als inhaltliche Grundlage, wobei der Aufbau einen großen Unterschied zu den bisherigen Prüfungen aufweist. Er beinhaltet zwar die per GMK-Beschluss festgelegten drei Prüfungsbereiche, diese wurden jedoch um jeweils einen ergänzt, sodass die gesamte Prüfung aus sechs Einzelteilen besteht. Im Bereich des Arzt\*-Patienten\*-Gesprächs erfolgt demnach neben der Anamnese ein OP-Aufklärungsgespräch. Der Kurz-Arztbrief zum Anamnese-Teil wird durch das Anfertigen des Teils Verlauf und Prozedere erweitert. Zudem ist bei dieser Prüfung sowohl ein Arzt\*-Arzt\*-Gespräch als auch ein Gespräch bzw. Telefonat mit einer/einem Angehörigen eines anderen Heilberufs (hier: Pflegepersonal) vorgesehen (vgl. Abb. 2).

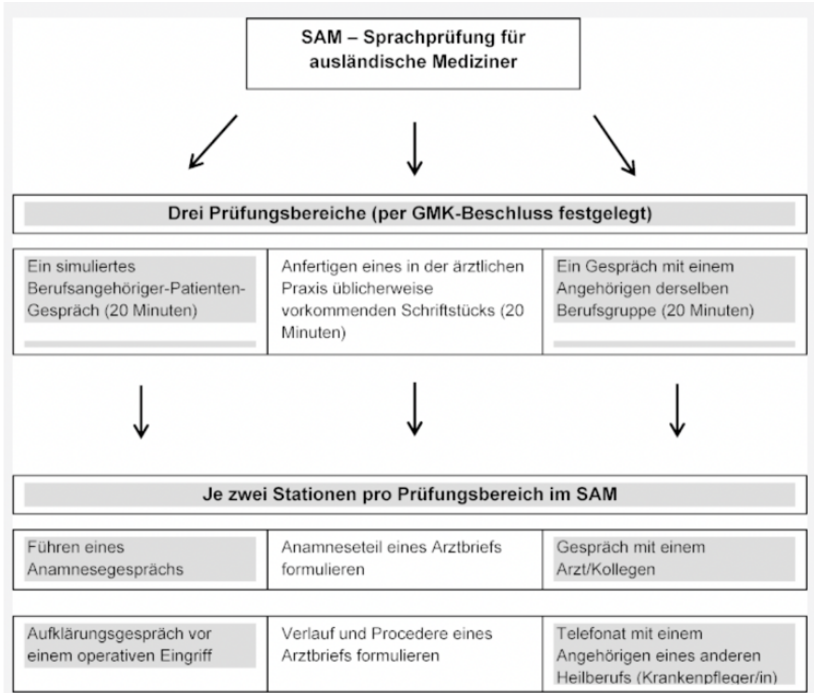


Abb. 2: Schematischer Aufbau des SAM unter der Berücksichtigung der GMK-Mindestanforderungen (Lenz et al. 2019: Abb. 1)

Für die vier Gespräche werden inklusive Vorbereitung je 10 Minuten und für die schriftlichen Teile nochmals 40 Minuten (2 x 20 Minuten) zur Durchführung angesetzt.

Die Fallvignetten für die vier mündlichen Teile des Tests stammen aus den Fachbereichen Allgemeinmedizin, Innere Medizin und Chirurgie. Die Wahl dieser Bereiche entspricht vorwiegend denen, die im 3. Staatsexamen vor Erhalt der Approbation sowie in der Kenntnisprüfung bei ausländischen Ärzt\*innen aus Drittstaaten integriert sind. Komplexe fachspezifische Inhalte werden laut Testautor\*innen bei den Fallszenarien vermieden, um die Prüfungsteile unabhängig von der fachlichen Spezialisierung der Prüfungsteilnehmenden bewältigen zu können (vgl. Lenz et al. 2019: o.S.). Anders als bei den anderen Prüfungen, erhalten die Teilnehmenden zu jedem Gespräch bzw.

bei jeder Station einen unterschiedlichen medizinischen Fall, was ebenso bei den zwei schriftlichen Teilen der Prüfung der Fall ist. Hier soll anhand einer auf Video gezeigten Anamnese ein Kurz-Arztbrief erstellt sowie auf Grundlage von kommentierten Laborberichten der Teilverlauf und Prozedere des Arztbriefes verfasst werden. Beide Teile entsprechen den i.d.R. vier typischen Strukturelementen eines Arztbriefes, die durch eine hohe (fach-)sprachliche Anforderung gekennzeichnet sind (vgl. ebd.). Bei diesen Teilen wird v.a. die „Rezeptionsfähigkeit und Verarbeitung sprachlichen Inputs sowie die schriftliche Ausdrucksfähigkeit“ (ebd.) geprüft.

Der Aufbau und die Zeitverteilung des SAM entsprechen dem im medizinischen Kontext international etablierten OSCE-Format (*Objective Structured Clinical Examination*), das als valides und reliables Instrument zur Überprüfung klinisch-praktischer Fähigkeiten gilt und sich darüber hinaus „für die Messung kommunikativer Kompetenzen in kulturellen und professionellen Kontexten eigne[t]“ (Bagnasco et al. (2016) zit. in: Lenz et al. 2019: o.S.)<sup>22</sup>. Die mündlichen Teilbereiche werden hierbei, im Gegensatz zu den anderen Prüfungen, per Video über eine testinterne Software aufgezeichnet. Dieses Verfahren gilt dann als sog. VOSCE (*Video-Recorded Objective Structured Clinical Examination*) und ist ein bewährtes Instrument zur Bewertung kommunikativer Fähigkeiten in medizinischen Kontexten (vgl. Lenz et al. 2019: o.S.). Bei dieser Prüfungs- und Bewertungsform übernehmen geschulte Schauspielpatient\*innen und externe Ärzt\*innen die Rolle der jeweiligen Gesprächspartner\*innen, wofür sie, neben einem Training, ein Skript mit Gesprächsanleitungen und Zusatzfragen erhalten. Die Prüfungsbewertung findet also nicht während der Prüfung durch die Anwesenden, sondern asynchron durch externe Bewerter\*innen statt. Dieses Vorgehen wird bspw. auch beim *TestDaF* angewandt.

Das Bewerter\*innen-Team besteht aus einer Ärztin/einem Arzt und einer/einem Sprachwissenschaftler\*in mit testmethodischem Fachwissen zu Deutsch

.....

22 Mehr zum OSCE-Format vgl. bspw. Nikendei C. & Jünger J. (2016): „OSCE – praktische Tipps zur Implementierung einer klinisch-praktischen Prüfung“, in: *GMS Z Med Ausbild.* 2006 23/3, (Doc 47), abrufbar unter: <http://www.egms.de/static/de/journals/zma/2006-23/zma000266.shtml>, 28.07.2020, Brandes H. (2006): *Überprüfung kommunikativer Fähigkeiten der Studierenden des Reformstudienganges Medizin der Charité Berlin mit Hilfe einer OSCE-Station*, Berlin: Charité Universitätsmedizin Berlin, Medizinischen Fakultät.

als Fremdsprache, die für die Pilotierung seitens der Testautor\*innen geschult wurden. Hierfür müssen pro Teilstation zwischen 11 und 17 Items bewertet werden, was einer Gesamtzahl von 83 Items entspricht.

Dem Bewerter\*innen-Team stehen dafür eine Bewertungsskala je Teil, als auch ein Beiblatt mit (sprachlichen und inhaltlichen) Ausführungen der Bewertungskriterien zur Verfügung. Die Skalen sind immer in drei Kategorien unterteilt, die sich je nach Prüfungsteil und sprachlicher/kommunikativer Handlung unterscheiden (vgl. Abb. 3).

Anamnese-Gespräch	Arztbrief: Verlauf und Prozedere
1) Erfassung der gesundheitliche Probleme	1) Zusammenfassen der Angaben
2) Kommunikatives Verhalten	2) Sprachliche Gestaltung
3) Gesamtbewertung	3) Gesamtbewertung
Beurteilung: jeweils mit <i>trifft eher zu</i> (1 Pkt.) – <i>uneindeutig</i> (0,5 Pkt.) – <i>trifft eher nicht zu</i> (0 Pkt.)	

Abb. 3: Bewertungskategorien am Beispiel für das *Anamnese-Gespräch* und den Teil *Arztbrief ‚Verlauf und Prozedere‘*

Die Bewertung erfolgt zunächst getrennt, anschließend werden die Ergebnisse verglichen, wobei sich das Bewerter\*innen-Team schließlich einstimmig über das Bestehen oder Nicht-Bestehen einigen muss. Dies erfordert eine Mindestpunktzahl pro Teilbereich. Die Gleichgewichtung aller Teilstationen soll dabei durch eine nachträgliche Anpassung<sup>23</sup> sichergestellt werden, damit die Itemanzahl nicht künstlich erhöht oder reduziert werden muss.

Eine im kleinen Rahmen stattgefundene Pilotierung an der LMU ergab gute Ergebnisse v.a. hinsichtlich der Test-Objektivität und Authentizität. Eine weitere Validierung wird in Form einer Vergleichsstudie des in Bayern eingesetzten Verfahrens durch die Ärztekammer und des SAM von unabhängigen Expert\*innen in Berlin durchgeführt.

.....

23 Die Anpassung erfolgt indem „die Punktzahl der sechs Teilstationen durch eine Multiplikation mit unterschiedlichen Faktoren so gewichtet werden, dass jeweils 1/6 der maximal möglichen Gesamtpunktzahl in jeder Teilstation erworben werden kann“ (Lenz et al. 2019: o.S.).

## 2.2 Analyse und Evaluation

Der Vergleich und die Analyse der vorgestellten Prüfungen haben gezeigt, dass sich alle Formate eng an den GMK-Beschlüssen orientieren. Alle überprüfen fachsprachliche Kenntnisse in mündlicher wie schriftlicher Form und weisen die dort geforderten fachspezifischen Kommunikationssituationen auf: Es finden ein Arzt\*-Patienten\*-Gespräch (Anamnese) sowie ein Arzt\*-Arzt\*-Gespräch (Patienten\*vorstellung) statt und eine berufstypische schriftliche Dokumentation (Arztbrief) wird verlangt. Angesichts des hierfür notwendigen Sprachniveaus kann von der (theoretischen) Überprüfung der Fachsprachenkenntnisse auf dem geforderten C1-Niveau bei allen gesprochen werden. Wie bereits in Kapitel 2.1 in Grundzügen dargestellt, unterscheiden sich die Prüfungen nichtsdestotrotz in gewissen Punkten erheblich. Diese möchte ich nun im Einzelnen spezifizieren und evaluieren.

### 2.2.1 *telc Deutsch B2·C1 Medizin Fachsprachprüfung*

Diese Prüfung ist die einzige der FSP, die alle vier Fertigkeiten (Hören, Lesen, Schreiben, Sprechen) explizit abprüft. Die Teile ‚Hörverstehen‘, ‚Leseverstehen‘ und ‚Sprachbausteine‘ werden in Form einer schriftlichen Gruppenprüfung vor der 60-minütigen Einzelprüfung abgelegt. Dieses Format bietet den Vorteil, beim Antrag auf Approbation bzw. Berufserlaubnis kein gesondertes B2-Zertifikat vorlegen zu müssen, da bei dieser Prüfung sowohl allgemeinsprachliche Kenntnisse auf B2 als auch Fachsprachenkenntnisse auf C1-Niveau abgeprüft werden (vgl. auch Marburger Bund 2018). Dies kommt dem in den GMK-Beschlüssen formulierten Ziel nach, „dass von den Antragstellenden für die Erteilung der Approbation oder Berufserlaubnis künftig [nicht] verpflichtend zwei Sprachtests gefordert werden [sollen]“ (87. GMK 2014: 2). Das Augenmerk der Analyse lag auf der Einzelprüfung, die den Inhalten und Zeitvorgaben der GMK-Beschlüsse entspricht. Für die Anamnese bekommen die Teilnehmenden fünf Minuten Vorbereitungszeit und zu keinem Zeitpunkt dürfen Hilfsmittel genutzt werden. Für die dort beschriebenen drei Kommunikationssituationen kommt bei dieser *telc*-Prüfung ein Fall zum Einsatz. Das bedeutet die Prüfungsteilnehmenden bekommen zur Durchführung aller Teile einen Patienten\*-Fall, der ein bestimmtes Krankheitsbild mit den entsprechenden

Symptomen usw. darstellt. Dieser wird beim Anamnese-Gespräch erhoben, dazu der Kurz-Arztbrief verfasst sowie beim Arzt\*-Arzt\*-Gespräch vorgestellt. Diese Vorgehensweise ist in der Hinsicht problematisch, dass dadurch die Bewertungskriterien der Prüfungsbereiche in Abhängigkeit zu einander stehen und somit „[d]ie Leistung in einem Bereich [...] nicht mehr ausschließlich von der Kompetenz in diesem Bereich ab[hängt], sondern auch von der Leistung in bereits durchlaufenen Testabschnitten“ (Moosbrugger & Kelava 2008, zit. nach Lenz et al. 2019: o.S.). Darüber hinaus entsteht dabei das Risiko, je nach Fall bzw. dessen Fachgebiet, Prüfungsteilnehmenden durch ihre bisherige berufliche Erfahrung zu bevorteilen bzw. zu benachteiligen. Denn besseres Fachwissen führt automatisch auch zur besseren Prüfungsleitung, auch wenn bei der FSP das medizinische Fachwissen per se nicht zu bewerten gilt. Die Prüfungsfairness könnte somit gefährdet oder zumindest reduziert sein. Die Tatsache, dass der schriftliche Teil der Einzelprüfung (Teil 2: Kurz-Arztbrief) im selben Raum in Anwesenheit der Prüfenden verfasst wird, könnte zur erhöhten Nervosität führen. Ein weiterer problematischer Punkt stellt die synchrone Bewertung der mündlichen Prüfungsleistungen dar. Wie bereits beschrieben, führen die Prüfung zwei Prüfer\*innen bzw. Ärzt\*innen nach vorgegebenen Bewertungskriterien durch. Die Beteiligung der bewertenden Person am Prüfungsgespräch erschwert dabei jedoch einen objektiven Bewertungsfokus. Ferner ist das Gesagte immer flüchtig und eine Wiederholung ist nicht möglich (vgl. ebd.). Positiv zu bewerten ist dabei, dass sowohl die/der *telc*-Prüfende als auch die Ärztin/der Arzt eine *telc*-Lizenzierung benötigen und somit speziell für dieses Prüfungsformat geschult worden sind<sup>24</sup>. Die Bewertungskriterien dieser FSP entsprechen denen einer ‚4-Fertigkeits-Prüfung‘, die sowohl sprachliche als auch inhaltliche Angemessenheit berücksichtigen und wird dem Format somit gerecht. Es kann jedoch gerade bei der mündlichen Prüfung diskutiert werden, ob die allgemeine Einteilung bzw. Gewichtung von *inhaltlicher Aufgabenbewältigung*, *Aussprache/Intonation*, *Flüssigkeit*, *Korrektheit* und *Wortschatz* hinsichtlich des Ziels einer berufsspezifischen ‚Sprachhandlungskompetenz‘ (vgl. Kapitel 2.1) zulänglich ist. Für die zwei mündlichen Teile der Einzelprüfung (Teil 1 und 3) gibt es drei Bewertungsbereiche, da nach der Fallvorstel-

.....

24 Zur Problematik dieser Prüfer\*qualifizierungen siehe den Beitrag von Grein in diesem Band.

lung noch (mindestens drei) fallbezogene Rückfragen seitens der Ärztin/des Arztes gestellt werden, die mit bewertet werden. Fallbezogene Fragen bergen das Risiko, Fachwissen abzufragen und u.U. schlechtere sprachliche Leistungen aufgrund fehlender Fachkenntnisse zu erbringen. Insgesamt wird hier jeweils zwischen B2 und C1: gut erfüllt oder erfüllt (B1 oder unter B1) entschieden (vgl. *telc* Übungstest 1: 47). Die Prüfer\*innen müssen sich nicht auf ein gemeinsames Ergebnis einigen, der gerundete Mittelwert ermittelt das Endergebnis (vgl. ebd.). Dies erfolgt später durch lizenzierte Bewerter\*innen in der *telc* Zentrale und kann bis zur Übermittlung der Ergebnisse bzw. Ausstellung des Zertifikats einige Wochen dauern. Dies kann angesichts der Abhängigkeit der Ergebnisse für den weiteren Verlauf des Anerkennungsverfahrens sowie den damit verbundenen Aufenthaltsstatus (bei Nicht-EU-Staatsbürger\*innen) nicht unproblematisch sein.

Neben der hier dargestellten *telc Deutsch B2·C1 Medizin Fachsprachprüfung* werden auch die *telc Deutsch B2 Medizin Zugangsprüfung* und die skalierte Prüfung *telc Deutsch B2·C1 Medizin*<sup>25</sup>, angeboten. Es bleibt fraglich, inwiefern ein tatsächlicher Bedarf an drei verschiedenen Prüfungsformaten im Bereich Medizin notwendig und fachlich nachvollziehbar ist, zumal sie fachbereichsunspezifisch sind. Positiv zu sehen sind die 3.000 *telc*-Prüfungszentren weltweit<sup>26</sup>, die ein flexibleres Ablegen der Prüfung ermöglichen, wobei sich die Frage stellt, an wie vielen dieser Zentren v.a. im Ausland auch die FSP Medizin tatsächlich durchgeführt wird. Außerdem bietet die *telc* Zentrale in Frankfurt a.M. zentrale Prüfungstermine an (wegen Corona vorerst ausgesetzt). Zudem steht ein kostenloser Übungstest zum Download zur Verfügung. Dies unterstützt eine gute Vorbereitungsmöglichkeit für die Teilnehmenden und Kursleitenden und schafft damit Transparenz nicht nur hinsichtlich der Prüfungsteile und des Ablaufs, sondern auch der Bewertung(skriterien). Die *telc gGmbH* als

.....

25 Skalierte Prüfungen umfassen zwei unterschiedliche GER-Stufen. Die Aufgaben in der Prüfung sind für zwei aufeinanderfolgende Niveaus entwickelt und erlaubt eine Auswertung der Sprachkompetenz auf zwei unterschiedlichen GER-Stufen. Für weiterführende Informationen zu diesen Prüfungsformaten siehe: <https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/filter.html>, 30.07.2020.

26 Für eine Suche bzw. Übersicht der weltweiten *telc*-Prüfungszentren vgl. <https://www.telc.net/pruefungsteilnehmende/pruefungszentrum-finden.html>, 30.07.2020.



etablierter und erfahrener Prüfungsanbieter hat sich den *ALTE*-Standards der Testentwicklung verschrieben. Über statistische Auswertungen, den Teilnahmezahlen, Bestehens- und Durchfallquote gibt es keine zugänglichen Daten.

Diese FSP wird allerdings nur in drei Bundesländern (Hessen, Saarland, Schleswig-Holstein) als erforderlicher Sprachnachweis bei der Approbations-Beantragung akzeptiert und dürfte daher anderorts bei den Antragsstellenden kaum sehr attraktiv sein. Diese Regelung fördert ferner die Problematik des o.g. Prüfungstourismus.

### Zusammenfassende Darstellung *telc Deutsch B2-C1 Medizin Fachsprachprüfung*

<b>Aufbau:</b>	B2-Niveau: alltagsprachlich, C1-Niveau: fachsprachlich; Schriftliche Gruppenprüfung und Einzelprüfung: Anamnese – Arztbrief – Fallvorstellung
<b>Teile + Dauer</b>	Schriftliche Gruppenprüfung: 80 Min. [HV: 30Min., LV: 40 Min., Sprachbausteine: 10Min.]; Einzelprüfung: 65 Min.= je Teil 20 Min. (+ 5 Min. Vorbereitung)
<b>Patientenfälle</b>	1
<b>Bewerter*innen</b>	2: mind. 1 Ärztin/Arzt und 1 <i>telc</i> -lizenzierte Prüfer*in beide: Prüferlizenz
<b>Schulung/Lizenz</b>	1-tägige Schulung zum Prüfungsformat durch <i>telc</i> -Referent*in
<b>Gesprächspartner*innen für</b>	
<b>Anamnese</b>	Patient*in = prüfende Ärztin/Arzt
<b>Fallvorstellung</b>	Ärztin/Arzt = prüfende Ärztin/Arzt
<b>Bewertung</b>	
<b>mündlich</b>	synchron durch Prüfer*innen, keine Einigung nötig → Endnote = Mittelwert (durch <i>telc</i> -Zentrale bestimmt)
<b>schriftlich</b>	durch <i>telc</i> -Bewerter*in mit Bewerterlizenz
<b>Punkte und Kriterien</b>	mündlich: 15 Punkte – inhaltliche Angemessenheit; 20 Punkte: sprachliche Angemessenheit sowie Aussprache/Intonation, Flüssigkeit, Korrektheit, und Wortschatz Arztbrief: 15 Punkte inhaltliche; Angemessenheit/Aufgabenbewältigung, Kommunikative Gestaltung, Korrektheit und Wortschatz Verteilung: Schriftliche Gruppenprüfung: 50% – 50% Einzelprüfung

bestanden mit	60% je Teil (mündlich/schriftlich);
Wiederholung bei Nicht-Bestehen	Gesamtprüfung, beliebig oft
Anforderungen	Anmeldung und Gebühr
Material (zur Vorbereitung)	Modelltest inkl. Hördateien, Bewertungsbögen, Ablaufplan
Konzeption	interne Testentwicklung nach ALTE-Standards
durchgeführt von	telc-Zentrale oder lizenziertes Testzentrum
anerkannt in	Hessen, Saarland, Schleswig-Holstein
Kosten	abhängig vom jeweiligen Testzentrum
digitale Version	X

### 2.2.2 Ärztliche Fachsprachenprüfung Rheinland-Pfalz

Diese FSP weist neben den im GMK-Beschluss formulierten inhaltlichen und zeitlichen Mindestanforderung auch weitere Prüfungsteile bzw. -abschnitte auf (vgl. Kapitel 2.1.2). Es werden gesondert zusätzlich das Hör- und Leseverstehen getestet sowie medizinische Fachbegriffe in Form eines Vokabeltests und Beschriftungen eines Körperschemas abgefragt. Von einer umfangreichen ‚4-Fertigkeiten-Prüfung‘ kann hierbei jedoch nicht gesprochen werden. Seit Juni bzw. Juli 2020 existieren aktualisierte Informationen zur FSP in RLP und ein dazugehöriges Informationsblatt mit ‚Ablaufhinweisen‘ (vgl. BÄK 2020a/b). Daraus geht zum einen die Änderung des ersten Teils vom bisherigen ‚Vokabeltest‘ zum ‚Verständnistest‘ und damit die Erweiterung der Unteraufgabe c) *Beschriftungen einer Schemazeichnung des Körpers* hervor. Außerdem wurde als Teil 2 das *freie Gespräch* eingeführt. Die weiteren Teile wurden beibehalten, wobei Teil 4 und 5 zusammengefügt zu sein scheinen. Innerhalb der Informationsblätter herrscht also keine Übereinstimmung der Prüfungsteile und Ausführungen. Außerdem sind keine konkreten Erläuterungen oder Anforderungen für Teil 2 *freies Gespräch*, bei Teil 3 zur *Patienten\*aufklärung* und zu *Anforderung zu Untersuchungen* zu finden. Laut Aussagen der Prüfungsteilnehmenden handelt es sich beim *freien Gespräch* um eine Vorstellung der eigenen Person. Mit der *Patienten\*aufklärung* ist das an das Anamnesegespräch anschließende Informieren der Patientin/des Patienten über das weitere Vorgehen bzw. weitere Untersuchungen gemeint. Bei letzterem Abschnitt müssen zwei Anforderungs-

scheine z.B. für eine Computertomografie (CT) mit fehlenden Informationen wie bspw. den Grund der durchzuführenden Untersuchung in Form eines Lückentextes ausgefüllt werden. Diese können allerdings, im Gegensatz zu den Ausführungen auf dem Informationsblatt, unabhängig von der durchgeführten Anamnese und des vorliegenden Patienten\*falls seitens der Prüfenden gewählt werden. Dass diese wichtigen Informationen nicht explizit aufgeführt sind, zeugt von fehlender Transparenz seitens der Bezirksärztekammer und mindert eine gute Vorbereitung.

Mit dem Einsetzen eines Vokabeltests, wie es bei dieser FSP der Fall ist, werden die Leistungen isoliert erfasst, was eine verlässliche und eindeutig ‚falsch‘ oder ‚richtig‘ zu bewertende Lösung mit sich bringt (vgl. Rossa 2016: 400). Damit lässt sich zwar eine höhere Reliabilität erzeugen, es fehlt aber die Kompetenzorientierung, da die so wichtige Authentizität der Prüfungs- bzw. Sprachleistung damit verloren geht. Die Effektivität des isolierten Abfragens bzw. ‚Übersetzens‘ von Fachvokabular in Laiensprache und umgekehrt<sup>27</sup> könnte man hinsichtlich des Ziels einer ‚Sprachhandlungskompetenz‘ (vgl. Kapitel 1.1) im medizinischen Kontext als Prüfungsteil grundsätzlich in Frage stellen, zumal diese Kompetenz im Rahmen des Patienten\*-/Arzt\*-Arzt\*-Gesprächs überprüft werden kann bzw. wird. Wie bei der *telc*-Prüfung kommt auch hier die ‚Ein-Fall-Szenario‘-Methode bei den mündlichen Teilen zum Einsatz. Über die oben genannten Problematiken hinaus, die damit einhergehen, dass sich nur ein einziger Patienten\*fall über mehrere Prüfungsfälle erstreckt, ist bei dieser Methode der Prüfungsverschleiß deutlich höher. Der Austausch eines Falls bedeutet automatisch den Austausch des gesamten Tests oder zumindest der aufeinander aufbauenden Teile. Die Schwierigkeit des neuen Falls lässt sich somit nicht eindeutig vergleichen. Laut Berichten seitens der Prüfungsteilnehmenden sind ca. 22 abwechselnd eingesetzte Krankheits-/Patientenfälle

.....

27 Das Übersetzen/Dolmetschen (interlinguale Mediation) und das Paraphrasieren in der Fremdsprache (intra-linguale Mediation) wurden zwar im neuen *Begleitband des GER* aufgenommen, doch entspricht dieser Vokabeltest nicht dem Konzept der Sprachmittlung (vgl. *Begleitband zum GER*: Englische Version abrufbar unter: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>, 11.08.2020; für Deutsch vgl. Quetz, J. & Rossa, H. (Hrsg.) (2019): „Der neue Companion Volume to the Common European Framework of Reference for Languages (Begleitband zum Gemeinsamen europäischen Referenzrahmen für Sprachen)“ in: *Zeitschrift für Fremdsprachenforschung* 30: 2.

bekannt, die anscheinend kürzlich um weitere sieben ergänzt wurden (bspw. auch Covid-19).

Beim Abschnitt zum Leseverstehen sollen zu ärztlichen Befunden bzw. einem Entlassungsbrief drei Fragen schriftlich auf dem dafür ausgehändigten Vordruck beantwortet werden. In Bezug auf die Höraufgabe in diesem Teil herrschen jedoch Unklarheiten: Von zwei Telefonanrufen sollen Informationen notiert werden. Dabei ist nicht klar, wer anruft und in welcher Form die Dokumentation erfolgen soll. Dass es sich dabei um Anrufe bspw. aus dem Labor zur Übermittlung von Patienten\*informationen handelt, erfährt man nur über die TN.

Wie in Kapitel 2.1.2 beschrieben, müssen die jeweiligen Teile mit einer „ihm zugeordneten“ Mindestpunktzahl abgeschlossen werden. Das Nichterreichen dieser führt zum direkten Abbruch der Prüfung und die weiteren Prüfungsteile werden nicht durchgeführt. Über die erforderlichen Mindestpunkte existieren allerdings keine zugänglichen Informationen. Zu den Prüfungsteilen heißt es zudem, dass sie in unterschiedlicher Abfolge geprüft werden können (BÄK 2020a: 2). Dies erscheint hinsichtlich der ‚Abbruchsregel‘ höchst problematisch und kann zu hohem Verlust der Prüfungsfairness führen sowie den Fortlauf der Prüfung massiv beeinflussen, zumal die Wertigkeit bzw. die Gewichtung der Abschnitte unterschiedlich ist. Die Bewertung findet hierbei, wie bei *telc*, synchron statt. Dazu übernehmen auch hier die Prüfenden gleichzeitig die Rolle der/des Bewertenden und der/des Gesprächspartner\*in. Bei der ärzttekammerlichen Prüfung besteht die Prüfungskommission jedoch lediglich aus Ärzt\*innen. Die Aussage von Dr. Windhorst, Präsident der Ärztekammer Westfalen-Lippe, die als unabhängige Prüfungsinstanz die Durchführung der FSP für das Land durchführt, untermauert diese Wahl:

Wer besser als eine Ärztin oder ein Arzt kann beurteilen, ob medizinisch Sprachinhalte ausreichend sicher verstanden und so weitergegeben werden können, dass Missverständnisse vermieden werden, Patientinnen und Patient sprachlich empathisch begegnet werden kann und ärztliche Kolleginnen und Kollegen aus den Angaben die medizinisch richtigen Schlussfolgerungen ziehen können (zit. in: ÄKWL 2019: o.S.).

Dieser Ansicht steht die Forderung gegenüber, angesichts der Komplexität der Fachsprachenprüfung beide Fachrichtungen (Medizin und Sprachwissenschaft) zur Bewertung der Prüfungsteilnehmenden heranzuziehen (vgl. Deppe et al. 2015: 37 und Kapitel 2.3), zumal nicht bekannt ist, ob die prüfenden Ärzt\*innen eine Art Prüfungsschulung (in Bezug auf sprachwissenschaftliche Kriterien) durchlaufen (müssen). Im Zusammenhang mit einer validen Fachsprachenprüfung spielen unweigerlich die Prüfer\*innen, v.a. die die mündlichen Sprachfertigkeiten der Ärzt\*innen bewerten, eine ausschlaggebende Rolle (vgl. Kapitel 2.3). Es stellt sich also die Frage, welche Kompetenzen diese zur Abnahme von berufs-/fachspezifischen Prüfungen und zur Bewertung der (mündlichen) Sprachkompetenzen benötigen und wie sie diese erlangen. Wie bereits beschrieben, erfahren die Prüfungsteilnehmenden direkt im Anschluss an die Prüfung über ihr Bestehen und bekommen ggf. ein Dokument zum Nachweis ausgehändigt. Im Falle des Nichtbestehens wird keine Bescheinigung ausgestellt, was im Rahmen der BAMF-geförderten Maßnahme ‚Spezialmodul: Heilberufe‘ insofern problematisch ist, da somit kein Nachweis für die Teilnahme an der FSP vorgelegt werden und sich dies ggf. negativ auf die weitere (Kurs)Förderung auswirken kann.

Zur Vorbereitung auf die Prüfung kann auf der Internetseite der Ärztekammer der Anamnesebogen, der für die Erhebung in Teil 3 und zur Ausformulierung in Teil 4 bereitgestellt wird, genutzt werden; ein Modelltest der vollständigen Prüfung existiert nicht. Die Anfrage seitens eines Kursträgers in Mainz, ob die Kursleitungen der BAMF-geförderten ‚Mediziner-Kurse‘ bei der FSP in Mainz hospitieren dürften, um im Unterricht eine bessere Prüfungsvorbereitung zu gewährleisten, wurde ohne handfeste Begründung abgelehnt. Dies war vor einigen Jahren noch möglich.

Hinsichtlich der ausführlichen Beschreibungen und Berichte der abgelegten Prüfung (z.B. Auflistung aller Fachbegriffe zu Teil 1 oder Darstellung der Patient\*innen-Anamnese), die im Internet gerne ausgetauscht und zur Verfügung gestellt werden, kommt die Frage auf, inwiefern die Prüfungsunterlagen seitens der TN so im Detail dokumentiert werden können, da sie nicht mitgenommen werden dürfen. Nichtsdestotrotz liefern genau diese Informationen mehr Aufschluss über den konkreten Ablauf und geben eine Art ‚Innenansicht‘ dieser Prüfung, die sonst hinsichtlich der Konzeption, der Gewichtung und

Mindestpunktezah der Prüfungsteile, der Bewertung(sriterien), der Materialauswahl und der Prüfer\*innenqualifikation bzw. -kommission einige Fragen offenlässt.

Auf dieser Informationsgrundlage lässt sich ein wissenschaftlich fundiertes Konzept dieser Prüfung nur schwer erkennen. Ein umfänglicherer Datenzugang bzw. die Auskunft seitens der Bezirksärztekammer ist für eine umfassendere und aussagekräftigere Analyse und Beurteilung notwendig. Aber genau dieser eher intransparente und unwissenschaftliche ‚Ist-Zustand‘ bildet die aktuelle Lage der FSP ab, nicht nur in Rheinland-Pfalz, sondern die der meisten Ärztekammern in Deutschland (vgl. bspw. Malakhova & Lenz 2018).

### Zusammenfassende Darstellung FSP in RLP

<b>Aufbau:</b>	5 Teile; können in unterschiedlicher Abfolge geprüft werden, bei Nicht-Erreichen der Mindestpunktzahl des jeweiligen Teils → sofortiger Abbruch der Prüfung
<b>Teile + Dauer</b>	1) Verständnistest 2) Freies Gespräch 3) Anamnese/Aufklärung 4) Kurz-Arztbrief 5) Lesen + Hören und Patientenvorstellung; Teil 3, 4 und 5 (Fallvorstellung): je 20 Min., Rest unbekannt → insgesamt 2,5–3 Std. inkl. Wartezeit
<b>Patientenfälle</b>	1
<b>Bewerter*innen</b>	2: approbierte Ärzt*innen, Teil der Prüfungskommission der Ärztekammer
<b>Schulung/Lizenz</b>	Nicht bekannt
<b>Gesprächspartner*innen für</b>	
<b>Anamnese</b>	Patient*in = prüfende Ärztin/Arzt
<b>Fallvorstellung</b>	Ärztin/Arzt = prüfende Ärztin/Arzt
<b>Bewertung</b>	
<b>mündlich</b>	synchron durch Prüfer*innen
<b>schriftlich</b>	synchron durch Prüfer*innen
<b>Punkte und Kriterien</b>	<ul style="list-style-type: none"> <li>• Mindestpunktzahl je Teil (bei Nicht-Erreichen → Abbruch der Prüfung)</li> <li>• unterschiedliche Wertigkeit je Teil, Details unbekannt</li> </ul>
<b>bestanden mit</b>	Gesamtpunktzahl aus allen Prüfungsteilen: 60%

<b>Wiederholung bei Nicht-Bestehen</b>	Gesamtprüfung, beliebig oft
<b>Anforderungen</b>	Prüfungsantrag: – kurzer Lebenslauf, – Kopie der qualifizierten Eingangsbestätigung (= Beantragung auf Erteilung einer Berufserlaubnis/Approbation in RLP) – Kopie eines Sprachzertifikates, falls vorhanden: Beleg über ein Job-Angebot in Rheinland-Pfalz / über eine laufende oder abgeschlossene Hospitation in Rheinland-Pfalz; Gebühr
<b>Material (zur Vorbereitung)</b>	Ablauf, Hinweise zu Prüfungsteilen, Anamnesebogen (als Download)
<b>Konzeption</b>	Nicht bekannt
<b>durchgeführt von</b>	Bezirksärztekammer RLP
<b>anerkannt in</b>	RLP
<b>Kosten</b>	425€ (Stand 2020)
<b>digitale Version</b>	X

### 2.2.3 SAM – *Sprachtest für ausländische Mediziner*

Beim SAM handelt es sich um eine reine Fachsprachenprüfung, die ebenfalls rezeptive und produktive Sprachleistungen testet. Dies geschieht in realitätsnahen und authentischen Kommunikationssituationen und schließt als einziger Test auch die Überprüfung der (fach)sprachlichen Kompetenzen mit Angehörigen anderer Heilberufe mit ein. Dieser Gesprächsbereich ist in den GMK-Beschlüssen ebenfalls beschrieben (vgl. 87.GMK 2014: 3) und wird im SAM in Form eines Anweisungsgesprächs (Telefonat) mit einem/einer Krankenpfleger\*in umgesetzt (vgl. Abb. 2). Dieses sowie die Ergänzung eines Aufklärungsgesprächs vor einer Operation und das Verfassen eines zweiten Teils des Arztbriefs stellen weitere berufstypische Situationen bei der ärztlichen Tätigkeit im Krankenhaus oder in einer Klinik dar. Die insgesamt sechs unterschiedlichen Teile (mündlich wie schriftlich) decken somit nicht nur einen größeren realitätsnahen Umfang ab, sondern ermöglichen auch einen geringeren Prüfungsverschleiß. Die voneinander unabhängigen Prüfungsteile erleichtern den Austausch älterer Fallszenarien und ein größerer Pool an Prüfungsversionen erhöht die Prüfungssicherheit. Die Vermeidung der sonst eingesetzten ‚Ein-Fall-Methode‘ wirkt sich darüber hinaus positiv auf die Test-Fairness aus.

Die Authentizität der Prüfungssituation spiegelt sich auch in der Zeitvorgabe von 10 Minuten pro Station wider. Die sehr kurz erscheinende Prüfungszeit entspricht jedoch einer realistischen Abbildung der zeitlichen Ressourcen im Arbeitsalltag von Ärzt\*innen (vgl. Lenz et al. 2019: o.S.). Für eine Patientenvorstellung gibt es keine 20 Minuten Zeit und muss binnen weniger Minuten präzise übermittelt werden. Auch für eine Anamnese kann ein/e (Haus)Ärzt\*in sich nicht so viel Zeit nehmen<sup>28</sup>. Diese Kompetenz wird im Beruf auch von ausländischen Ärzt\*innen verlangt und beim SAM auf diese Weise abgeprüft. Die Aufteilung in mehrere, aber dafür kürzere Prüfungsgespräche bzw. -stationen entspricht dem für die mündlichen Kommunikationssituationen gewählte OSCE-Format (vgl. Kapitel 3.1.3). Das Verhalten der Prüfungsteilnehmenden kann dadurch insgesamt viermal in unterschiedlichen Kontexten beobachtet werden und erhöht somit die Reliabilität.

Bei der Wahl der Prüfungs- und Bewertungsform als auch bei der Wahl der Aufgaben wurde auf wissenschaftliche Kriterien geachtet. Beim VOSCE handelt es sich um eine erprobte Methode zur Bewertung kommunikativer Fähigkeiten in medizinischen Kontexten. Die Erstellung der Bewertungskriterien, sprich die einzelnen Items der Bewertungsskala sowie die Ausführungen zur Beurteilung der Sprachkompetenz, wurden für jede Teilstation auf Basis der Analyse wissenschaftlicher Literatur sowie empirischer Beispiele entwickelt (vgl. Lenz et al. 2019: o.S.). Die Bewertung enthält keine klassischen Bewertungskategorien, wie sie bspw. bei *telc* standardmäßig eingesetzt werden, sondern beurteilen viel mehr die sprachliche Handlungskompetenz. Neben der Fachsprachlichkeit wird hier insgesamt eine situationsangemessene Sprachleistung bewertet, die neben Sozialkompetenzen (z.B. *Der Arzt/die Ärztin spricht den Patienten/die Patientin respektvoll an.*) auch interkulturelle Elemente wie ‚mündige/r Patient\*in‘, Begrüßungsformen und Hörerrückmeldung enthalten. Das Einbeziehen dieser Elemente, die nicht im GMK-Beschluss verankert sind, zeugt von der Sinnhaftigkeit einer interdisziplinären Konzeption der Prüfung unter Berücksichtigung von Sprachwissenschaftler\*innen. Die asynchrone Bewertung eines exter-

.....

28 Laut einer Statistik werden bei einem Ländervergleich hierfür in Deutschland ca. 7,5 Minuten genannt (vgl. Figure 2 in: Irving, G. et al. (2017).



nen Bewerter\*innen-Teams, das nicht am Prüfungsgespräch selbst beteiligt ist, lässt sich v.a. hinsichtlich der Objektivität sehr positiv bewerten. Darüber hinaus besteht das Team hierbei immer aus einer Ärztin/einem Arzt und einer/einem Sprachwissenschaftler\*in mit Deutsch als Fremdsprache-Hintergrund, wodurch beide vertretenden Fachbereiche einbezogen werden. Die Einführung und Etablierung einer standardisierten Bewerter\*innenschulung ist hierbei stark zu empfehlen, um eine reliable und objektive Bewertung gewährleisten zu können (vgl. Kapitel 2.3). Das gleiche gilt bezüglich der für die Durchführung eingesetzten Schauspielpatient\*innen bzw. Ärzt\*innen, wobei hier bereits ein ausgebautes Schulungsangebot (z.B. bereits während des Medizin-Studiums) existiert und somit auf geschulte und erfahrene Personen zurückgegriffen werden kann. Für die Durchführung der einzelnen Stationen entsteht aufgrund des Prüfungsformats ein erhöhter Bedarf an benötigten Räumlichkeiten. Diese Aspekte werden beim SAM seitens der Testentwickler\*innen selbst als „Schwäche“ bezeichnet, da für die Prüfungslogistik besonders zu Beginn einige Investitionen (testinterne Aufnahmesoftware, Schulung der Schauspielpatient\*innen und der Bewerter\*innenteams) geleistet werden müssen. Langfristig „sollten die Stärken des SAM diesen Nachteil allerdings mehr als ausgleichen“ (Lenz et al. 2019: o.S.). Diese FSP weist nach der Auswertung der ersten Pilotierung eine vielversprechend hohe Test-Objektivität und Skalengültigkeit auf, die in authentischen Kommunikationssituationen fair und objektiv die fachsprachlichen Anforderungen auf C1-Niveau überprüft. Eine Pilotierung im großen Umfang und weitere Erprobungen sind sicherlich notwendig, nicht zuletzt, um eine gewünschte Erhöhung der Reliabilität zu erreichen und sicherstellen zu können.

## Zusammenfassende Darstellung SAM

<b>Aufbau:</b>	6 Stationen nach OSCE-Format, 2 pro GMK-Bereich
<b>Teile + Dauer</b>	1) Anamnese 2) OP-Aufklärung 3) Arztbrief: Anamnese 4) Arztbrief: Verlauf und Prozedere 5) Fallvorstellung 6) Telefonat mit Krankenpfleger*in; 4 x 10 Min. (mündlich) + 2 x 20 Min. (schriftlich) = 80 Min.
<b>Patientenfälle</b>	6, je Station ein anderer
<b>Bewerter*innen</b>	2: 1 Ärztin/Arzt und 1 Sprachwissenschaftler*in/DaF-ler*in
<b>Schulung/Lizenz</b>	Schulung zu fallspezifischen Sprachhandlungen → pro Station: Gesprächsanleitungen + Zusatzfragen
<b>Gesprächspartner*innen für</b>	
<b>Anamnese</b>	Patient*in/Pflegekraft = geschulte Schauspieler*in
<b>Fallvorstellung</b>	Ärztin/Arzt = externe Ärztin/externer Arzt
<b>Bewertung</b>	
<b>mündlich</b>	asynchrone Bewertung durch Bewerber*innenteam (anhand von Videoaufzeichnungen)
<b>schriftlich</b>	asynchrone Bewertung durch Bewerber*innenteam
<b>Punkte und Kriterien</b>	<ul style="list-style-type: none"> <li>• pro Teilstation: Bewertung zwischen 11 und 17 Items → Gesamtzahl: 83 Items (nachträgliche Anpassung zur Gleichgewichtung) Bewertungsskala: je drei Kategorien → Beurteilung: je mit trifft eher zu (1 Pkt.) – uneindeutig (0,5 Pkt.) – trifft eher nicht zu (0 Pkt.)</li> <li>• Beiblatt mit (sprachlichen und inhaltlichen) Ausführungen der Bewertungskriterien je Station</li> <li>• getrennte Bewertung → Einstimmigkeit bei Endergebnis (Bestehen oder Nicht-Bestehen)</li> </ul>
<b>bestanden mit</b>	kumulative Leistung aller sechs Teilbereichen: mind. 60%
<b>Wiederholung bei Nicht-Bestehen</b>	Noch nicht bekannt
<b>Anforderungen</b>	B2-Zertifikat
<b>Material (zur Vorbereitung)</b>	noch nicht vollumfänglich verfügbar, Bsp. für Bewertungsbogen und -kriterien, Ausführungen für Bewerber*innen-Team
<b>Konzeption</b>	interdisziplinär: Didaktik und Ausbildungsforschung in der Medizin, Deutsch als Fremdsprache, empirische Pädagogik und pädagogische Psychologie; wissenschaftlich fundiert

durchgeführt von	in der Etablierungsphase
anerkannt in	in der Etablierungsphase
Kosten	Noch nicht bekannt
digitale Version	in der Planungsphase

### 3 Zusammenfassender Vergleich und Diskussion

Die vorgestellten Prüfungsformate stellen alle eine Fachsprachenprüfung nach Vorgaben des GMK-Beschlusses dar, die medizinspezifische Sprachkenntnisse (auf C1-Niveau) überprüfen.

Der inhaltliche Fokus aller drei Prüfungen liegt auf der medizinischen Kommunikation im Krankenhaus oder in einer Klinik und den dortigen Gesprächssituationen bzw. -konstellationen (z.B. Bericht an Stationsarzt\*in – Oberarzt\*in, mit dem Pflegepersonal, OP-Aufklärung). Angesichts des erhöhten Mangels an Hausarzt\*innen v.a. in ländlichen Regionen (vgl. bspw. Lübking 2020 und BPB 2012) werden diese den typischen Kommunikationssituationen in einer Praxis nicht gerecht. Lenz argumentiert hingegen, dass in einer hausärztlichen Praxis der Kommunikationsumfang eingeschränkter sei, meistens würden dort nur Anamnesegespräche geführt, was für eine ärztliche FSP unzulänglich wäre (vgl. Interview 2020). Außerdem sollte einerseits ein/e approbierte/r Ärztin/Arzt all solche Gesprächssituationen meistern können. Andererseits stellt die FSP eine Überprüfung allgemeiner medizinischer Fachsprachenkompetenzen dar, mit deren Nachweis man auch in einem Krankenhaus in Deutschland tätig werden kann. Eine Anpassung der Kurs- und Prüfungsformate ist m.E. hinsichtlich einer allgemeinen Unterscheidung der Zielgruppen im Kontext Medizin und der speziellen (Sprach)Bedarfe fachlich in Erwägung zu ziehen. Diese Empfehlung wurde sowohl seitens des deutschen Klinikpersonals in einer Studie von Soppa (vgl. 2019: 59), als auch seitens von Prüfungsteilnehmenden ausgesprochen<sup>29</sup>. Eine praktische Umsetzung scheint aufgrund des hohen (Kosten)Aufwands jedoch kaum realistisch.

.....  
 29 Vgl. Umfrage im Beitrag von Jehle in diesem Band.

Nennenswerte Unterschiede zwischen den Prüfungen zeichnen sich v.a. in den Bereichen Inhalte und Aufbau, Bewertungsform und -kriterien sowie Prüfer\*innen ab, was unmittelbaren Einfluss auf die testtheoretische Güte (vgl. Kapitel 2) der FSP hat.

Die von den Ärztekammern durchgeführte FSP ist bislang in den meisten Bundesländern die einzig anerkannte FSP, daher stellen deren Unterschiede Prüfungsteilnehmenden vor unnötige Herausforderungen und ergibt m.E. einen umso dringlicheren Bedarf an Einheitlichkeit und Transparenz. Die Erfüllung der von der GMK beschlossenen Mindestanforderungen sind hierfür jedoch nicht ausreichend: Die Vorgaben zum konkreten Ablauf sind zu vage und konkrete Bewertungsrichtlinien sind nicht vorhanden. Dies bietet den Ärztekammern noch zu viel Spielraum bei der Durchführung (vgl. auch Malakhova & Lenz 2018: 3). Der SAM, den es hier mit etablierten FSP zu vergleichen galt, bietet einen guten Ansatz hin zu einem bundeseinheitlichen Prüfungssystem, das mit einem wissenschaftlich fundierten Konzept den aufgezeigten Problemen (vgl. Kapitel 2.3) entgegenwirkt.

Insgesamt kann das Ziel der Repräsentativität und der Möglichkeit einer Generalisierung der Ergebnisse auf die Gesamtheit der Prüfungen mithilfe dieser Analyse nicht zufriedenstellend realisiert werden. Die Ergebnisse ermöglichen jedoch einen Einblick auf die aktuelle Prüfungswelt und können so als Basis für weitere Forschungsvorhaben dienen.

Mit der Einführung eines neuen Prüfungsformats sollte auch eine Anpassung bzw. Erweiterung der Sprachkurse für Mediziner\*innen einhergehen, da nur bei bekanntem Testform auch dessen Testgüte greift. Hierbei müsste zwischen Kursen zur reinen Testvorbereitung und/oder als Sprachlernangebot unterschieden werden. Roche betont, dass gerade berufsspezifische Sprachkurse „[...] stärker als bisher fallbasiert und szenarien-didaktisch arbeiten [sollten], also stärker die tatsächliche Kommunikation in den Blick nehmen als die abstrakten Regeln des Sprachsystems. Auch die Sprachprüfungen müssten diese Kompetenzen stärker berücksichtigen“ (2014: 318). Dies entspricht auch dem grundsätzlichen Ziel des ‚Masterplan Medizinstudium 2020‘<sup>30</sup>, mit dem in

.....  
30 Vgl. ‚Masterplan Medizinstudium 2020‘, abrufbar unter: <https://www.bmbf.de/de/masterplan-medizinstudium-2020-4024.html>, 31.07.2020.

der Approbationsverordnung ein stärkerer Fokus auf Kommunikationskompetenzen sowie vermehrt kompetenzorientierte Studieninhalte und (Prüfungs-) Aufgaben verankert sind (statt des bisherigen ‚Ankreuzens‘). So wird auch bei den in Deutschland ausgebildeten Ärzt\*innen eine insgesamt praxis- und handlungsorientiertere Lehre stattfinden.

Im Zuge einer solchen Etablierung wäre ebenso eine Weiterbildung der KL hinsichtlich der Anforderungs- und Bewertungskriterien des neuen Prüfungsformats ratsam und notwendig, um die internationalen Ärzt\*innen bestmöglich auf die Prüfung vorbereiten zu können. Im Fall des SAM wäre das Bereitstellen von Beispielvideos der einzelnen Prüfungsteile inklusive Bewertungskriterien für KL und TN sehr hilfreich.

Der Personalmangel im medizinischen Bereich sollte und darf nie zu einem Herabsetzen der zur Berufsausübung fachlichen wie sprachlichen Anforderungen führen, auch nicht zu Berufsbeginn. Berufe, bei denen die Patient\*innensicherheit oberstes Gebot darstellt und die Gesundheit auf dem Spiel stehen kann, bergen durchaus beachtliche Risiken. Eine Einheitlichkeit der FSP in den Bundesländern bei Durchführung und Bewertung wäre hierfür sehr förderlich. Grundlegend ist darüber hinaus die Einstellung der internationalen Ärzt\*innen, die sich nicht ausschließlich auf ihre fachlichen Kenntnisse und medizinischen Fähigkeiten verlassen dürfen, sondern der Sprache und ihrer kompetenten Verwendung (auf allgemein- und fachsprachlicher Ebene) höhere Wichtigkeit beimessen müssen. Dies ist letztlich für das Bestehen der FSP ausschlaggebend.

Der Blick in die Praxis zeigt, dass die für eine reibungslose Berufsausübung erforderlichen (Fach)Sprachkompetenzen trotz bestandener FSP nicht immer ausreichend vorhanden sind (vgl. bspw. Studie Soppa 2019) und darüber hinaus weitere Faktoren für den erfolgreichen Arbeitsalltag eine wichtige Rolle spielen (vgl. ebd.). Diese komplexen (Sprach)Kompetenzen vermag keine FSP in vollem Umfang zu überprüfen, sodass über ein kompetenzerweitertes Anerkennungsverfahren bzw. damit verbundene Fördermaßnahmen zu diskutieren wäre. Neben einer validen FSP, die als Nachweis der für den Berufsalltag unabdingbaren (Fach)Sprachkenntnisse gilt, sollten darüber hinaus auch die ebenso wichtigen interkulturellen Kompetenzen gefördert werden, zum einen durch eine verstärkte Einbindung der Thematik ‚Kultursensible Medizin‘ und

„Interkulturelle Kommunikation“ im Sprachunterricht. Zum anderen könnte bspw. seitens der medizinischen Einrichtung vor Arbeitsbeginn ein Interkulturelles Training für ausländische Ärzt\*innen verpflichtend angeboten als auch berufsbegleitende (Sprach)Förderangebote etabliert werden. Zusätzlich sollte auch das deutsche Klinikpersonal im Rahmen kontinuierlicher Angebote nicht nur im Umgang mit Patient\*innen mit Migrationshintergrund geschult, sondern auch für die kulturellen und arbeitstechnischen Unterschiede in einem internationalen Team sensibilisiert werden. So kann der Berufseinstieg und eine langfristige Integration ausländischer Ärzt\*innen erleichtert werden.

Angesichts des großen Bedarfs an Pflegekräften v.a. in der Altenpflege (vgl. Augurzyk & Kolodziej 2018: o.S.) ist die Forderung nach einem bundesweit einheitlichen Anerkennungsverfahren sowie der fachspezifischen sprachlichen Handlungsfähigkeit und deren Nachweis auch auf diesen Bereich übertragbar bzw. auf alle (verkammerten und unverkammerten) Heilberufe, wie Zahnärzt\*innen und Apotheker\*innen.

Es bleibt also spannend, wie sich die Anerkennungsverfahren bzw. die Prüfungslandschaft im Bereich der FSP Medizin sowie der gesamte Integrationsprozess und die Förderangebote ausländischer Fachkräfte in Zukunft entwickeln werden. Bei einem föderalen System wie dem hiesigen, in dem immer viel von politischen Entscheidungen abhängt und v.a. in diesem Bereich mitunter auch wirtschaftlich entschieden wird, ist eine flächendeckende Einführung einer „neuen“ Prüfung schwer umsetzbar. Nichtsdestotrotz bietet der SAM aus dargelegten Gründen die Voraussetzungen dafür, „ein Teil des Approbationsverfahrens ausländischer [Ärzt\*innen]“ (Lenz et al. 2019: o.S.) zu werden. Dies ist bspw. in Form eines weiteren anerkannten Anbieters (neben der FSP der Landesärztekammer) im jeweiligen Bundesland möglich und hätte so die Chance auf Etablierung. Doch solange es eine Auswahl an und bedeutende Unterschiede der anerkannten Prüfungen gibt sowie seitens der Prüfungsteilnehmenden der Weg des geringsten Aufwands und Widerstands gewählt wird, stellt sich die Frage, inwieweit der aus Teilnehmendensicht komplex und umfangreicher scheinende SAM<sup>31</sup> freiwillig gewählt würde. Geringere Gebühren könnten dabei sicherlich einen Anreiz schaffen.

.....  
31 Vgl. Umfrage im Beitrag von Jehle in diesem Band.

Eine denkbare digitale Version dieses Prüfungsformats bietet in Zeiten von Digitalisierung und Globalisierung – nicht nur angesichts der Corona-Pandemie – weitere Möglichkeiten der weltweiten und v.a. ressourcensparenden Prüfungsdurchführung.

## 4 Fazit

Der vorliegende Beitrag sollte zum einen zeigen, ob eine FSP als Nachweis der von internationalen Ärzt\*innen geforderten Sprachkompetenzen gelten kann, zum anderen inwiefern die vorgestellten FSP einen validen Test zur Überprüfung dieser darstellen. Die erste Fragestellung kann hinsichtlich der komplexen und vielschichten (sprachlichen) Anforderungen im ärztlichen Tätigkeitsfeld mit ‚nein‘ beantwortet werden. Die beschriebene ‚sprachliche Handlungskompetenz‘ kann im Rahmen einer FSP nicht vollumfänglich überprüft werden, v.a. angesichts der derzeitigen Prüfungslandschaft. Voraussetzung für die Aussagekraft des Nachweises ist eine Standardisierung der Tests und ihrer Bewertung. Hier hat der Vergleich der Prüfungen gezeigt, dass der SAM einen wissenschaftlich fundierten Ansatz liefert, der mit seiner Etablierung einen Beitrag zur einheitlichen und validen Überprüfung fachspezifischer Sprachkompetenzen leisten könnte. Zudem wurde deutlich, wie sinnvoll und notwendig eine interdisziplinäre Konzeption und Bewertung von FSP sind. Nur unter Einbeziehen aller tangierender Fachbereiche ist die Gewährleistung der Testgüte und somit eines validen Nachweises möglich. Nicht zuletzt soll dieser Beitrag für ein kompetenzorientiertes Anerkennungsverfahren sensibilisieren.

## Bibliografie

ALTE (2012): *Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER*, erstellt von ALTE im Auftrag des Europarats/Abteilung für Sprachenpolitik, Frankfurt: telc gGmbH, abrufbar unter: [https://www.telc.net/fileadmin/user\\_upload/handbuch\\_zur\\_entwicklung\\_und\\_durchfuehrung\\_von\\_sprachtests.pdf](https://www.telc.net/fileadmin/user_upload/handbuch_zur_entwicklung_und_durchfuehrung_von_sprachtests.pdf), 25.07.2020.

- AUGURZKY, B. & KOLODZIEJ, I. (2018): *Fachkräftebedarf im Gesundheits- und Sozialwesen 2030. Gutachten im Auftrag des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung*, RWI – Leibniz-Institut für Wirtschaftsforschung, abrufbar unter: [https://www.sachverstaendigenrat-wirtschaft.de/fileadmin/dateiablage/gutachten/jg201819/arbeitspapiere/Arbeitspapier\\_06-2018.pdf](https://www.sachverstaendigenrat-wirtschaft.de/fileadmin/dateiablage/gutachten/jg201819/arbeitspapiere/Arbeitspapier_06-2018.pdf), 25.07.2020.
- BAGNASCO A. et al. (2016): „How to maintain equity and objectivity in assessing the communication skills in a large group of student nurses during a long examination session, using the Objective Structured Clinical Examination (OSCE)“, in: *Nurse Educ Today* 38: 54–60.
- BOROWSKI, D. (2018): *Sprachliche Herausforderungen ausländischer Anästhesist(inn)en bei Aufklärungsgesprächen. Eine gesprächsanalytische Studie zu „Deutsch als Zweitsprache im Beruf“*, Berlin: Frank & Timme.
- DEPPE, J. et al. (2015): *Praxisguide Deutsch im Krankenhaus. Die sprachliche Integration internationaler Ärztinnen und Ärzte. Eine Handreichung für Klinikleitungen, Deutschlehrkräfte und Bildungsfachleute*, Hamburg: Fachstelle Berufsbezogenes Deutsch im Förderprogramm IQ, abrufbar unter: [http://www.deutsch-am-arbeitsplatz.de/fileadmin/user\\_upload/PDF/Praxisguide\\_Web.pdf](http://www.deutsch-am-arbeitsplatz.de/fileadmin/user_upload/PDF/Praxisguide_Web.pdf), 27.07.2020.
- DIEKMANN, A. (2009): *Empirische Sozialforschung. Grundlagen Methoden Anwendungen*, Hamburg: Rowohlt.
- DIPPOLD-SCHENK, K. (2018): *Die Entwicklung eines didaktischen Designs im Rahmen einer modularen Weiterbildungsreihe zur Professionalisierung von Sprachlehrenden für den Bereich Deutsch-Medizin*. Dissertation, abrufbar unter: <https://opus4.kobv.de/opus4-fau/frontdoor/index/index/year/2018/docId/9735>, 20.07.2020
- HANDT, G. V. D. (2002): „Sprachtests und Zertifikate“, in: QUETZ, J. & HANDT, G. V. D. (Hrsg.): *Neue Sprachen lehren und lernen: Fremdsprachenunterricht in der Weiterbildung*, Bielefeld: Bertelsmann: 187–193.
- IRVING, G. et al. (2017): „International variations in primary care physician consultation time: a systematic review of 67 countries“, in: *BMJ Open* 7/10, abrufbar unter: <https://bmjopen.bmj.com/content/7/10/e017902>, 04.11.2020.



- JEHLE, N. (2019): „Vermittlung fachspezifischer Fremdsprachenkompetenzen am Beispiel des Pilotprojekts ‚Fachspezifischer Deutschkurs: Humanmedizin‘ für Erasmus-Studierende an der JGU Mainz“, in: ERSCH, C. M. (Hrsg.): *Kompetenzen in DaF/DaZ*, Berlin: Frank & Timme: 41–66, abrufbar unter: <https://library.oapen.org/handle/20.500.12657/24334>, 21.10.2020.
- KELLNHAUSER, E. & SCHEWIOR-POPP, S. (1999): *Ausländische Patienten besser verstehen*, Stuttgart: Thieme.
- LENZ, H. et al. (2019): „Sprache zählt: Entwicklung eines objektiven strukturierten Sprachtests für ausländische Ärztinnen und Ärzte – Ergebnisse einer Pilotstudie in Deutschland“, in: *GMS Journal for Medical Education* 36/1, (Doc 2), abrufbar unter: <https://www.egms.de/static/de/journals/zma/2019-36/zma001210.shtml#ref24>, 31.07.2020.
- LÜBKING, U. (2020): *Sicherstellung der Gesundheitsversorgung auf dem Lande*, abrufbar unter: [https://www.nomos-elibrary.de/10.5771/1611-5821-2020-2-8.pdf?download\\_full\\_pdf=1](https://www.nomos-elibrary.de/10.5771/1611-5821-2020-2-8.pdf?download_full_pdf=1), 21.09.2020.
- MALAKHOVA, O. & LENZ, H. (2018): *Fachsprachprüfung Medizin in Deutschland. Der Versuch einer ersten Bestandsaufnahme* (unveröffentlichtes Manuskript).
- MOOSBRUGGER H. & KELAVA A. (2008): *Testtheorie und Fragebogenkonstruktion*, Berlin: Springer, abrufbar unter: <https://link.springer.com/book/10.1007%2F978-3-540-71635-8>, 30.07.2020.
- PERLMANN-BALME, M. (2016): „Sprachzertifikate“, in: BURWITZ-MELZER, E. et al. (Hrsg.): *Handbuch Fremdsprachenunterricht*, Tübingen: Francke: 420–423.
- ROCHE, J. (2014): „Sprache und Beruf – Zur Frage der Deutschkenntnisse in medizinischen Berufen“, in: *Bayrisches Ärzteblatt* 6: 316–318.
- ROELCKE, T. (2005): *Fachsprachen*, Berlin: Erich Schmidt Verlag.
- ROSSA, H. (2016): „Testen und Prüfen“, in: BURWITZ-MELZER, E. et al. (Hrsg.): *Handbuch Fremdsprachenunterricht*, Tübingen: Francke: 399–403.
- SCHMELTER, L. (2014): „Gütekriterien“, in: DEMIRKAYA, S. et al. (Hrsg.): *Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung*, Paderborn: Ferdinand Schöningh: 33–45.

- SHAFFER, N. & WIEDENKELLER, E. (2012): „Test(aufgaben)entwicklung am Beispiel einer neuen standardisierten B1-Prüfung für DaF“, in: CLALÜNA, M. & TSCHARNER, B. (Hrsg.): *Beurteilen im DaF-/DaZ-Unterricht Testen – Evaluieren – Prüfen*, Akten der Vierten Gesamtschweizerischen Tagung für Deutschlehrerinnen und Deutschlehrer 29. und 30. Juni 2012 – Universität Bern: 107–116.
- SOPPA, R. (2019): *Ausländische Ärztinnen und Ärzte – Chancen und Herausforderungen im DaZ-Unterricht*, Masterarbeit zur Erlangung des Akademischen Grades, Johannes Gutenberg-Universität Mainz.

## Internetquellen

- ÄKWL (2019): Ärztekammer Westfalen-Lippe: *Heilkunst braucht Sprachkunst*: 5.000 Fachsprachenprüfungen durch die Ärztekammer Westfalen-Lippe abgenommen – Pionierarbeit auf dem Feld der berufsspezifischen Sprachprüfung, abrufbar unter: [https://www.aekwl.de/fileadmin/user\\_upload/aekwl/Presse/Pressemitteilungen/12\\_19\\_5000\\_FSP.pdf](https://www.aekwl.de/fileadmin/user_upload/aekwl/Presse/Pressemitteilungen/12_19_5000_FSP.pdf), 30.07.2020.
- BÄK (o.J.): Bezirksärztekammer Rheinhausen: *Über uns*, abrufbar unter: <https://aerztekammer-mainz.de/kammer.php>, 27.07.2020.
- BÄK (2020a): Bezirksärztekammer Rheinhausen: *Informationen zur Fachsprachenprüfung (FSP) für ausländische Ärzte in Rheinland-Pfalz*, abrufbar unter: [https://aerztekammer-mainz.de/pdf/diverse/FSP-Informationen-2020\\_deutsch.pdf](https://aerztekammer-mainz.de/pdf/diverse/FSP-Informationen-2020_deutsch.pdf), 31.07.2020.
- BÄK (2020b): Bezirksärztekammer Rheinhausen: *Ablaufhinweise zur Fachsprachenprüfung (FSP) in Mainz*, abrufbar unter: <https://aerztekammer-mainz.de/pdf/diverse/FSP-Ablaufhinweise-2020.pdf>, 27.07.2020.
- BPB (2012): Bundeszentrale für politische Bildung: *Medizinische Versorgung im ländlichen Raum*, abrufbar unter: <https://www.bpb.de/politik/innenpolitik/gesundheitspolitik/169804/medizinische-versorgung-im-laendlichen-raum>, 21.09.2020.
- GMK: 87. Gesundheitsministerkonferenz am 26./27. Juni 2014 in Hamburg, TOP 7: *3 Eckpunkte zur Überprüfung der für die Berufsausübung erforderlichen Sprachkenntnisse*, abrufbar unter: [https://www.gesundheitsministerium.de/SharedDocs/Pressemitteilungen/DE/2014/06/20140626\\_gmk\\_top7.html](https://www.gesundheitsministerium.de/SharedDocs/Pressemitteilungen/DE/2014/06/20140626_gmk_top7.html), 27.07.2020.

*derlichen Deutschkenntnisse in den akademischen Heilberufen*, abrufbar unter: [https://www.gmkonline.de/documents/TOP73BerichtP\\_Oeffentl\\_Bereich.pdf](https://www.gmkonline.de/documents/TOP73BerichtP_Oeffentl_Bereich.pdf), 26.07.2020.

KMK (2018): Kultusministerkonferenz: *Handreichungen*, abrufbar unter: [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2011/2011\\_09\\_23-GEP-Handreichung.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2011/2011_09_23-GEP-Handreichung.pdf), 31.07.2020.

MARBURGER BUND (2018): *Deutschkenntnisse – Anforderungen in den Bundesländern für die Approbationserteilung*, abrufbar unter: <https://www.marburger-bund.de/sites/default/files/files/2018-09/deutschkenntnisse-german-requirements-approbation.pdf>, 28.07.2020.

TELC: *telc Deutsch B2·C1 Medizin Fachsprachprüfung*, Übungstest 1, abrufbar unter: <https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/detail/telc-deutsch-b2c1-medizin-fachsprachpruefung.html#t=2>, 27.07.2020.

TELC-a: *Wer sind wir*, abrufbar unter: <https://www.telc.net/ueber-telc/wer-wir-sind.html>, 19.10.2020.


TELC-b: *Prüfungsübersicht: Pflege und Medizin*, abrufbar unter: <https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/filter.html>, 19.10.2020.

TELC-c: *telc Deutsch B2·C1 Medizin Fachsprachprüfung*, abrufbar unter: [https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/download.html?type=5555&tx\\_telclangtest\\_render%5B-content%5D=3549&cHash=5da2bb41139ddfc44de552ac3a0c7d20](https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/download.html?type=5555&tx_telclangtest_render%5B-content%5D=3549&cHash=5da2bb41139ddfc44de552ac3a0c7d20), 19.10.2020.

## Anhang: Prüfungsformat Fachsprachenprüfung Medizin

(telc Deutsch B2-C1 Medizin Fachsprachprüfung, abrufbar unter: [https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/download.html?type=5555&tx\\_telclangtest\\_render%5Bcontent%5D=3549&cHash=5da2bb41139ddfc44de552ac3a0c7d20](https://www.telc.net/pruefungsteilnehmende/sprachpruefungen/pruefungen/download.html?type=5555&tx_telclangtest_render%5Bcontent%5D=3549&cHash=5da2bb41139ddfc44de552ac3a0c7d20), 19.10.2020.)

### Format der Prüfung

	Subtest	Ziel	Aufgabentyp	Zeit	Punkte
Schriftliche Gruppenprüfung	 <b>Hörverstehen</b>				
	Teil 1	Routinemäßige Situation verstehen	6 Richtig/Falsch- und 3 Mehrfachwahlaufgaben	30 Min.	20
	Teil 2	Unterschiedliche Positionen zu einer Fachthematik erkennen	5 Zuordnungsaufgaben		
	Teil 3	Fachdiskussion verfolgen	6 Richtig/Falsch-Aufgaben		
	 <b>Leseverstehen</b>				
	Teil 1	In einem Zeitungsartikel Informationen suchen und zuordnen	5 Mehrfachwahlaufgaben	40 Min.	20
	Teil 2	E-Mails verstehen und Betreffzeilen zuordnen	3 Zuordnungsaufgaben		
	Teil 3	Informationstext zu Regel- und Verfahrensabläufen verstehen	3 Richtig/Falsch- und 3 Mehrfachwahlaufgaben		
	Teil 4	In einem Internet-Fachforum Informationen suchen und zuordnen	6 Zuordnungsaufgaben		
	 <b>Sprachbausteine</b>				
	Passende diskurssteuernde Elemente in Routinegesprächen auswählen	10 Zuordnungsaufgaben	10 Min.	10	
Gesamtdauer der Schriftlichen Gruppenprüfung				80 Min.	
Einzelprüfung	<b>Vorbereitung</b>		Aufgabenblatt für Teil 1	5 Min.	
	 <b>Mündlicher Ausdruck</b>				
	Teil 1	Arzt-Patienten-Gespräch	Aufgabenblatt (Information über Patienten)	20 Min.	35'
	 <b>Schriftlicher Ausdruck</b>				
	Teil 2	Verfassen eines Kurz-Arztbriefes	Anamnesegespräch und -bogen	20 Min.	15
 <b>Mündlicher Ausdruck</b>					
Teil 3	Fallpräsentation/Fallvorstellung Rückfragen zur Präsentation	Anamnesegespräch und -bogen	20 Min.	35'	
Gesamtdauer der Einzelprüfung				85 Min.	



## Prüferqualifizierung – Fokus auf Mündlichkeit

### Einleitung

Mit der zunehmenden Globalisierung, den Internationalisierungsbestrebungen und dem ebenfalls ansteigenden Interesse, in Deutschland zu studieren, rückt das Thema „Prüfung“ zum Nachweis ausreichender Deutschkenntnisse immer mehr ins Zentrum des Interesses (vgl. Fandrych & Müller 2019: 293). Roche (2013: 108) macht mit seinem Statement „eine voreilige Diagnose von sprachlichen Fertigkeiten ohne die wissenschaftliche Fundierung solcher Tests kann verhängnisvoll werden“ deutlich, wie wichtig eine zuverlässige Bewertung von sprachlichen Kompetenzen ist. Während bis in die 90er Jahre die PNdS (Prüfung zum Nachweis der deutschen Sprachkenntnisse) quasi die einzige von Hochschulen durchgeführte Prüfung darstellte und lediglich zwischen bestanden und nicht bestanden differenzierte, entwickelten sich in den Folgejahren die *DSH* (*Deutsche Sprachprüfung für den Hochschulzugang*) und an den Goethe-Instituten das Große und das Kleine Deutsche Sprachdiplom (entwickelt an der LMU München) (vgl. Althaus 2018). 2001 kam der *TestDaF* hinzu (Test Deutsch als Fremdsprache), der sich durch einen hohen Grad an Standardisierung und weltweite Verbreitung in lizenzierten Testzentren von der *DSH* abgrenzte und weiterhin abgrenzt. 2004 versuchte die Hochschulrektorenkonferenz (HRK) eine stärkere Vereinheitlichung auch der *DSH*, 2019 wurde vom FaDaF und der HRK die *DSH*-Musterprüfungsordnung (MPO) überarbeitet (vgl. *DSH* Rundbrief No. 6); bereits 2004 wurde jedoch die aus der PNdS zunächst übernommene Dichotomie bestanden – nicht-bestanden zugunsten von abgrenzbaren Stufen abgeschafft (vgl. u.a. Koreik 2005, Demmig 2012). Im Jahre 2006, vollkommen unabhängig vom Hochschulmarkt, entwickelte die *telc gmbH* gemeinsam mit dem Goethe-Institut den Deutsch-Test für Zuwanderer (DTZ). *Telc* entwickelte dann zunächst eine C1-Prüfung, die

jedoch keine Anerkennung von der HRK und der Kultusministerkonferenz (KMK) erhielt. Diese wurde dann variiert als *telc C1 Hochschule* nach längerem Prozess in die Rahmenordnung über Deutsche Sprachprüfungen für das Studium an deutschen Hochschulen (RO-DT) aufgenommen (vgl. zu den Prüfungsformaten und Zulassungsvoraussetzungen Grein 2019; Altmeyer 2018). Die Universität Mainz beispielsweise nennt als *DSH-2 Äquivalente* (Quelle: <https://www.studium.uni-mainz.de/dsh-und-dsh-aequivalente/> (19.10. 2020)):

- das deutsche Abitur
- das Deutsche Sprachdiplom der Kultusministerkonferenz – Stufe II (DSD II)
- das Goethe-Zertifikat C2: Großes Deutsches Sprachdiplom (GDS)
- das Österreichische Sprachdiplom (ÖSD) – C1
- der *TestDaF* mit dem Ergebnis 4 x 4 oder besser
- die Feststellungsprüfung eines deutschen Studienkollegs
- das Zertifikat *telc Deutsch C1 Hochschule* ab dem Ergebnis „befriedigend“. (Hinweis: *telc Deutsch C1* ist nicht ausreichend, es muss sich um die Hochschulprüfung handeln).

Ziel des vorliegenden Beitrags ist es nun aber nicht, eine Validierung der einzelnen Sprachprüfungen vorzunehmen (siehe dazu Kecker 2011; Möhring & Bärenfänger 2018 zu Lese- und Wortschatzanforderungen *DSH*, *TestDaF* und *telc C1 Hochschule*), sondern zu erheben, welche Voraussetzungen die Prüferinnen und Prüfer haben müssen, um Prüfungen abzunehmen. Über welche Kompetenzen müssen also die „Prüfenden“ (oder auch Bewertenden) verfügen? Wie wird sichergestellt, vor allem im Bereich der mündlichen Prüfungen, dass die Prüfenden eine valide Einstufung vornehmen können? Wie erlangt man eine sog. Prüferlizenz? Berücksichtigt werden hier der *TestDaF*, *telc C1 Hochschule* und die *DSH*, da diese spezifische Sprachprüfungen für den Hochschulzugang anbieten. Nach einer Analyse der vorhandenen Dokumente werden Befragungen mit sowohl den Institutionen als auch zugelassenen Prüfenden aller drei Testformate durchgeführt. Nach einem allgemeinen Überblick über die Qualitätssicherung, mit Fokus auf mündliche Prüfungen, werden die drei Prüfungen und die Bewerter- oder Prüferqualifizierung skizziert. Es

folgt dann die Präsentation einer kleinen qualitativen Studie, in der Prüfer und Prüferinnen zu ihrer Qualifizierung befragt werden. Gerade Prüferinnen und Prüfer im Bereich der mündlichen Prüfungen müssen qualifiziert ausgebildet sein, um sicherzustellen, dass sie die Fähigkeiten der zu Prüfenden auch valide überprüfen können. Sie entscheiden mit ihrer Bewertung nicht selten über die Zulassung zum Studium.

## 1 Qualitätssicherung bei den Prüfenden – Sekundärquellen und Beschreibungen

Trotz intensiver Recherche findet sich wenig Sekundärliteratur zu den Anforderungen an die Prüfenden im Bereich mündlicher Sprachkompetenz. Lennon (2000: 169) führt die zentralen Kriterien für die Evaluation mündlicher Textproduktionen auf. Die einzelnen Bewertungskriterien und auch die Schulung der Prüfenden werden nicht thematisiert. Hoffmann (2014: 145) fasst zusammen: „Die Beurteilung mündlicher Leistung erfolgt normalerweise auf drei Ebenen, der formalen, der inhaltlichen und der strategischen.“ Formal wird Phonetik, Intonation, Flüssigkeit, lexikalische und grammatische Korrektheit überprüft. Auf der inhaltlichen Ebene wird auf die Eigenständigkeit der Sprachproduktion, den Grad der Orientierung an Vorlagen und die Situations- und Adressatengemessenheit der Äußerung geachtet. Der kommunikationsstrategische Bereich, also der Einsatz von Strategien, wird meines Wissens bei mündlichen Prüfungen nicht berücksichtigt. In Anlehnung an Straub (2014) werden sieben Kommunikationsstrategien differenziert: Verzögerungsstrategien (u.a. Themenwechsel, Ignorieren, Füllwörter), Sprachwechselstrategien, Erschließungsstrategien (u.a. Zustimmung/Verneinen), Kooperationsstrategien (u.a. Bitte um Hilfestellung, Bitte um Wiederholung oder Verdeutlichung, Verständniskontrolle), Imitationsstrategien (z.B. Nachahmung), Abrufungsstrategien (Anwendung von *chunks*) und nonverbale Strategien (Mimik, Gestik). Eva Šrámková (2016: 227) thematisiert konkret die Bewertung mündlicher Kompetenzen im universitären Bereich (in Tschechien). Sie (2016: 231) geht auch auf die Schulung der Prüfenden ein und formuliert:



Die Bewerter nehmen sowohl an den theoretisch als auch praktisch orientierten methodischen Schulungen und Seminaren teil, die nicht nur von den tschechischen, sondern auch von den ausländischen Experten durchgeführt werden. Zudem werden einige mündliche Prüfungen auf Video aufgezeichnet und im Rahmen der einzelnen Sprachsektionen regelmäßige Benchmarking-Treffen organisiert, bei denen die Leistungen der Studenten analysiert und bewertet werden mit dem Ziel, die Qualität der standardisierten Prüfungen unseres Fremdsprachenzentrums weiterhin zu verbessern.

Das *Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER* erstellt von ALTE im Auftrag des Europarats/Abteilung für Sprachenpolitik (2012) geht in seinem fünftem Kapitel auch auf den Bereich der Bewertung von Prüfungsleistungen ein. Auf S. 49 wird angeführt, dass die Bewerter einer Schulung bedürfen. Dazu genauer (2012: 50):

Bei nicht fest angestellten, jedoch regelmäßig eingesetzten Auswerterrinnen und Auswertern bietet sich ein Beurteilungssystem an, das auf Parametern wie Genauigkeit, Reliabilität und Geschwindigkeit basiert. Auswerter, die nicht zur Zufriedenheit gearbeitet haben, werden ggf. nicht weiter beschäftigt oder neu geschult. Eine solche Schulung kann Teil eines Systems sein [...]. Auswerter, die immer wieder eingesetzt werden, müssen nicht an jedem Training teilnehmen. Eine Begutachtung ihrer Leistung [...] macht es einfacher zu entscheiden, ob ein Auswerter zu einer Schulung gebeten wird, eine Auffrischung benötigt oder ersetzt werden soll.

Auf S. 52 wird konkret auch auf die Bewerterqualifizierung eingegangen. Dort heißt es:

Ziel der Schulungen von Bewerberinnen und Bewertern sind einheitliche und genaue Bewertungen. Standardisierung bedeutet, dass alle Bewerter die angestrebten Standards anwenden. [...] Es kann zudem notwendig sein, Bewertern ihre bereits vertrauten Bewertungsraster „abzutrainieren“.

Der Qualifizierungsprozess umfasst insgesamt mehrere Schritte von der offenen Diskussion hin zu unabhängiger Bewertung von Teilnehmerleistungen aus dem betreffenden Test:

- gelenkte Diskussion eines Teilnehmerbeispiels, durch das die Bewerber/innen die jeweilige Kompetenzstufe verstehen lernen
- unabhängige Bewertung eines Beispiels, gefolgt von einem Vergleich mit der Musterbewertung und Diskussion über die Gründe für abweichende Bewertungen
- unabhängige Bewertung mehrerer Beispiele, um zu erkennen, wie nah die Bewerber nun den Musterbewertungen kommen.

Auf S. 56 wird auch konkret auf den Bereich „mündliche Prüfung“ eingegangen und formuliert: „Mündliche Bewertung ist sehr viel schwieriger zu kontrollieren, es sei denn, die Leistung wird aufgezeichnet. Hier sollten also Bewerberinnen und Bewerber vor der Prüfung besonders gründlich geschult und richtig eingeschätzt werden.“

Notwendig sind folglich, neben ausgesprochen profunden Kenntnissen des *GER*, eine gründliche Schulung einerseits und andererseits die Digitalisierung der mündlichen Prüfung. Vorgeschlagen wird zur Überprüfung der Bewertungskompetenz neuer Bewerber diesen auch bereits bewertete Prüfungen vorzulegen, um derart zu erkennen, ob die Einschätzungen zuverlässig sind.

Auch der *EQUALS*-Rahmen (2019) geht auf den Bereich „Bewertung“ (S. 24–27) ein und benennt die Kompetenzen, über die ein Prüfender verfügen sollte. Dazu gehören u.a. Kenntnisse

- der Grundprinzipien von und der Unterschiede zwischen Bewerten, Beurteilen und Prüfen
- der Unterschiede zwischen formativer, summativer und diagnostischer Beurteilung, und wo welche Art am besten eingesetzt wird
- der möglichen Zusammenhänge zwischen Lehr-/Lern- und Beurteilungsverfahren
- der Grundlagen, des Aufbaus und Klassifizierungssystems des *GER* sowie deren Bedeutung für und Auswirkungen auf die Beurteilung

- der Schlüsselkriterien für die Gestaltung von Sprachprüfungen, Beurteilungen und Bewertungen sowie deren entsprechende Anwendung
- der unterschiedlichen Arten des Lernens, die den verschiedenen GER-Niveaustufen entsprechen, wie zum Beispiel vermehrtes Augenmerk auf kognitive Fertigkeiten ab Niveau B2, und entsprechende Auswirkungen auf die Beurteilung
- der Grundsätze und Durchführungsmöglichkeiten von formativer Beurteilung
- der differenzierteren Unterschiede bei den GER-Plus-Niveaustufen, zum Beispiel zwischen der B1+ Kompetenz im Vergleich zu B1 und B2
- der Prüfungsarten und Beurteilungstechniken in Bezug auf ihre Reliabilität (Zuverlässigkeit) und Validität (Gültigkeit) für den vorgesehenen Zweck, High oder Low Stakes usw.
- des Konzepts der Validität und der Unterschiede zwischen verschiedenen Arten der Validität sowie deren Auswirkungen auf Prüfungsgestaltung und -durchführung.

Wissenschaftliche Publikationen im Bereich der Prüferqualifizierung sind rar. Zurückgegriffen werden muss daher im Folgenden auch auf Darstellungen auf Homepages, z.B. zu den einzelnen Qualifizierungen und veröffentlichten Stellenanzeigen, die die Voraussetzungen für die Tätigkeit als „Beurteiler/in“ aufführen und, sofern vorhanden, Beurteiler-Handbücher. Sehr präzise werden die Voraussetzungen für die Abnahme von Sprachprüfungen beim Luftfahrt Bundesamt (2019) formuliert. Hier liegen „Leitlinien für den Erwerb und den Erhalt einer Befähigung zur Abnahme von Sprachprüfungen gem. § 125 LuftPersV“ vor. Neben Schulungsmaßnahmen wird hier ein abgeschlossenes sprachliches Studium gefordert. Die Schulungsmaßnahmen umfassen Kenntnisse der Rating Skalen, drei Bewertungen von Audioaufnahmen, Übungen zum Notizenerstellen; Übung zum Erkennen sprachlicher Defizite, eine Einheit zu administrativen und rechtlichen Aspekten der Prüfertätigkeit und eine Schulung zum Prüferverhalten.

Im Folgenden werden die Informationen zu den Prüferqualifizierungen für die Prüfungen, die zu einem Studium an einer deutschen Hochschule berech-

tigen, dargestellt: der *TestDaF*, die *telc C1 Hochschul*-Prüfung und die *DSH*. Das Goethe-Institut fordert seit 2015 ein Prüferzertifikat, auf das hier nicht eingegangen wird, da die Prüfungen des Goethe-Instituts nicht spezifisch auf den Zugang zu einer Universität ausgerichtet sind.

### 1.1 *TestDaF* (<https://www.testdaf.de>)

Der *TestDaF* wird in 96 Ländern an lizenzierten, d.h. offiziell zugelassenen Testzentren angeboten. Die Prüfung wird vom *TestDaF*-Institut in Bochum nach strengen Qualitätsstandards zentral entwickelt und ausgewertet. Die Prüfungsleistungen werden von Beurteilerinnen und Beurteilern bewertet. Die Evaluation von Norris & Drackert (2017) bestätigt die hohen Qualitätsstandards. Es liegt ein umfangreiches, immer wieder aktualisiertes Test-DaF Handbuch für Testzentren vor (aktuell Stand 2018). Hier findet sich auch ein Abschnitt „Ihre Verantwortung als Prüfungsbeauftragter“ (S. 16ff.). Das Schulungskonzept besteht aus dem genannten Handbuch, Präsenzseminaren für Prüfungsbeauftragte, sog. E-Modulen, und Webinaren. E-Module sind dabei Präsentationen, die einen Einstieg bieten; die Webinare (virtuelle Sitzungen) dienen vor allem dem Austausch zwischen den Mitarbeiterinnen und Mitarbeiter des *TestDaF*-Instituts und den Prüfenden sowie den Prüfenden untereinander. Ende Oktober 2020 erfolgte zudem die erste weltweite Online-Prüfung.

Der Test besteht aus vier Prüfungsteilen, die einen Bezug zum Hochschulkontext haben und dabei nicht fachspezifisch ausgerichtet sind. Im Bereich Leseverstehen (60 Minuten) werden drei Lesetexte mit 30 Items verwendet; Hörverstehen (40 Minuten) besteht ebenfalls aus drei Hörtexten mit 25 Items; schriftlicher Ausdruck (60 Minuten) beinhaltet eine Aufgabe zum Schreiben, und der für diese Studie herangezogene Bereich der mündlichen Prüfung („mündlicher Ausdruck“; 35 Minuten) besteht aus insgesamt sieben Sprechaufgaben. Hier sollen die Teilnehmenden zeigen, dass sie in verschiedenen Situationen an der Hochschule sprachlich handeln können. Dies umfasst z.B. Gespräche unter Studierenden, Beschreibungen, Stellungnahmen oder Hypothesenbildungen. Die mündlichen Textproduktionen werden dabei digitalisiert, so dass eine Überprüfung zeitlich flexibel wiederholt werden kann. In der online-Version sind leichte Veränderungen geplant.

Für die Beurteilenden liegt ein ausführliches Beurteilungshandbuch vor (hier: *Beurteilungshandbuch. Hinweise für Beurteilerinnen und Beurteiler zur Bewertung der produktiven Prüfungsteile beim TestDaF* 2015). Im Gegensatz zu den mündlichen Prüfungen von *telc* und der *DSH* kommt das Format des *Simulated Oral Proficiency Interviews (SOPI)* zum Einsatz, d.h., dass die Aufgaben digital abgespielt und die mündlichen Beiträge des Prüflings aufgezeichnet werden. Kritisiert wird hier zuweilen, dass dies „monologisch“ sei, da kein authentischer Dialog stattfindet und dies für den „ungewohnten“ Prüfling schwierig sein könne (vgl. z.B. Stevener 2018: 33). Aufgewogen wird diese Kritik jedoch durch die Digitalisierung der Sprachproduktionen, die eine Überprüfung der Einschätzung von Prüfenden ermöglicht, und die konkrete Aufforderung an die Prüfenden, diese Situation zu berücksichtigen (siehe Beurteilungshandbuch 2015: 25, 27).

Die mündliche Prüfung wird dabei nach drei Kriterien beurteilt:

1. Wie wirkt die Äußerung als Ganzes auf eine Hörerin oder einen Hörer? (Gesamteindruck)
  - Sprechen Sie flüssig, klar, verständlich?
  - Machen Aussprache und/oder Intonation das Verstehen leicht oder schwer?
  - Sind Aufbau und Struktur der Äußerung klar zu erkennen? Kann man dem Gedankengang gut folgen?
2. Erfüllt die Antwort die Aufgabenstellung? (Umsetzung der Aufgabe)
  - Bezieht sich die Äußerung auf das gestellte Thema?
  - Werden alle Punkte der Aufgabe ausreichend berücksichtigt?
  - Passt die Äußerung und passen die einzelnen Sprechhandlungen zu der Aufgabe und der Situation?
3. Mit welchen sprachlichen Mitteln wird die Aufgabe gelöst? (Sprachliche Realisierung)
  - Ist die Äußerung (Register, Aufbau, Anfang und Ende der Äußerung) der Situation angemessen?
  - Wie differenziert und angemessen sind Wortschatz und Syntax?

- Erschweren Fehler das Verstehen, oder kann man trotz Fehlern die Äußerung gut verstehen?

Es liegt mit dem Beurteilungshandbuch ein genaues Beurteilungsraster für die drei Bereiche vor. Die Prüfenden sind ferner dazu angehalten ihre Einstufungen auf dem Bewertungsbogen kurz zu kommentieren. Vor jeder Prüfung erhalten die Prüfenden sogenannte Kalibrierungsunterlagen. Diese bestehen aus einem Erwartungshorizont (also welche Leistungen werden erwartet), konkreten Beispielen aus der Erprobungsphase (digital) und eine begründete Beurteilung dieser Leistungen von einer Expertenrunde. Die Prüfenden werden folglich vor jeder Prüfung spezifisch auf diese Prüfung und die erwartbaren Lösungen vorbereitet. Vorgeschlagen wird dabei zunächst eine Einstufung der Prüfungsleistungen durch die Prüfenden selbst, die dann mit der Expertenbewertung abgeglichen wird. Der Bewertungsprozess ist folglich stark standardisiert. Bei der abschließenden Einstufung werden verzerrende Faktoren (Tendenzen zur Strenge bzw. Milde und Zentraltendenz, Positionseffekt, Halo-Effekt, Kontaminationseffekt und Persönlichkeits- und Situationsfaktoren) berücksichtigt. Ausgewertet wird mit dem Verfahren der Multifacetten-Rasch-Analyse (vgl. dazu die Publikationen von Eckes).

Der/die Prüfende (Beurteilende) muss über ein einschlägiges Hochschulstudium, Erfahrungen in der Durchführung von Sprachprüfungen verfügen und Deutschkenntnisse mindestens auf C1-Niveau vorweisen. Im Jahresbericht 2016/17 S. 43ff. wird konkret auf die Schwierigkeiten bei der Beurteilung mündlicher Prüfungsleistungen und die Notwendigkeit von Beurteilungskriterien eingegangen. Empirisch untersucht wird auch der sog. Rater-Ratee Effekt, also die Tendenz der Urteilsverzerrung abhängig von der Position des Beurteilers zu der zu beurteilenden Person. Insgesamt zeichnet sich der *TestDaF* durch wissenschaftliche Studien zur Qualitätssicherung und -entwicklung aus, die sich durch die enge Zusammenarbeit mit der Ruhr-Universität Bochum erklären lassen. Prüfende oder nach *TestDaF*-Terminologie Prüfungsbeauftragte werden in Seminaren geschult bzw. haben in Seminaren auch die Gelegenheit zum Austausch mit Kollegen. Konkret implementiert ist für Prüfende das „Seminar für Prüfungsbeauftragte“; daneben werden meist 2-tägige Workshops, aber auch kürzere Webinare, zur Fortbildung, ebenfalls für die Prüfenden,

angeboten. Themen sind hier – wie zu erwarten – Testen und Evaluieren und Leistungsbeurteilung. Eines der häufiger angebotenen Workshops/Webinare trägt dabei den Titel „Mündliche Leistungen im *TestDaF* beurteilen“.

In einer Stellenanzeige 2020 werden für die Stelle eines Prüfungsbeauftragten u.a. folgende Voraussetzungen formuliert:

- ein einschlägiger Studienabschluss, z. B. Magister, Master oder Staatsexamen (Sek II) in DaF, Germanistik oder einer modernen Fremdsprache
- zwei bis drei Jahre Lehrerfahrung in DaF auf den *GER*-Niveaus B2 und C1 sowie Erfahrungen im Umgang mit der Zielgruppe des *TestDaF*, ausländischen Studienbewerberinnen und -bewerber
- idealerweise Erfahrungen mit standardisierten Sprachprüfungen (als Prüferin/Prüfer oder Beurteilerin/Beurteiler)
- Bereitschaft zur Teilnahme an Schulungen des *TestDaF*-Instituts für Beurteilerinnen und Beurteiler.

Bereits 2003 zeigte Eckes (2003: 50), dass bei den Rasch-Modellen zur Qualitätssicherung im Rahmen von *TestDaF*-Erprobungsprüfungen auch eine Beurteilerschulung inkludiert ist, die Prüfenden also gezielt geschult werden. Kecker (2016: 155), Mitarbeiterin des *TestDaF*-Instituts, wird konkreter und schreibt zu den Prüfenden:

Der mit der Prüfungsdurchführung betraute verantwortliche Prüfungsbeauftragte muss bestimmte Voraussetzungen hinsichtlich seiner Qualifikation erfüllen und die in einem Handbuch beschriebenen Sicherheits- und Durchführungsbestimmungen beachten. Diese werden in Schulungsseminaren (virtuell oder Präsenzveranstaltung) erläutert, zu denen Prüfungsbeauftragte regelmäßig eingeladen werden. Die Einhaltung der Vorgaben wird durch unangekündigte Inspektionsbesuche am Prüfungstag kontrolliert.

Direkt zur Bewertung der mündlichen Produktionen formuliert Kecker (2016: 156):

Für die Bewertung der produktiven Leistungen im TestDaF wird im TestDaF Institut jeweils ein Beurteiler eingesetzt. Die Beurteiler müssen über eine entsprechende Lehrerfahrung mit der Zielgruppe verfügen und vor dem ersten Einsatz ein Schulungsseminar absolviert haben. Zusätzlich müssen sie jedes Jahr an einer weiteren Schulung teilnehmen, die dazu dient, eine einheitliche Anwendung der Bewertungskriterien aufrechtzuerhalten. Zu diesem Zweck werden detaillierte Handreichungen zum Bewertungsverfahren und Musterbewertungen, sog. Benchmarks, verwendet (vgl. Arras 2007). Darüber hinaus werden zu jedem Testlauf ein Erwartungshorizont und Benchmarks an die Beurteilergruppe verschickt.

Den Dokumenten ist zu entnehmen, dass beim *TestDaF* sehr viel Wert auf die Kompetenzen der Prüfenden gelegt wird.

## 1.2 *Telc* (The European Language Certificates) <https://www.telc.net/>

Die *telc gmbH* bietet über 90 standardisierte Prüfungen in insgesamt zehn Sprachen an. Dabei verfügt *telc* über 20.000 lizenzierte Prüferinnen und Prüfer weltweit. Sie ist ein Tochterunternehmen des Deutschen Volkshochschul-Verbands e.V. Die *Prüfung telc Deutsch C1 Hochschule* prüft hochschulbezogene Deutschkenntnisse. Insgesamt werden hier vier bzw. fünf Prüfungsteile angeboten: Leseverstehen, inklusive Sprachbausteine (90 Minuten), Hörverstehen (40 Minuten), schriftlicher Ausdruck (70 Minuten) und mündlicher Ausdruck (als Gruppenprüfung, je nachdem ob zwei oder drei Teilnehmende 16 oder 24 Minuten plus Vorbereitungszeit). Im ersten Teil halten sich die TN gegenseitig eine Präsentation zu einem vorgegebenen Thema. Jeweils im Anschluss müssen die Präsentationen der anderen zusammengefasst und kommentiert werden. Dabei sollen auch Fragen formuliert und diskutiert werden. Im zweiten Teil erfolgt eine Diskussion, wobei Wert auf eine nachvollziehbare Argumentation gelegt wird. Die Beurteilung erfolgt nach den Kriterien Aufgabengerechtigkeit, Flüssigkeit, Repertoire, grammatische Richtigkeit und Aussprache und Intonation. Inhaltliche Angemessenheit wird für jeden Teil getrennt bewertet, die



anderen Kriterien für die Prüfung insgesamt. Ein genaues Bewertungsschema liegt dabei vor (Bewertungsbogen M10) (vgl. Handbuch Deutsch Hochschule C1 2014). Ein konkreter Lizenzierungskurs für C1 Hochschule stand nicht zur Verfügung.

Zeidler (2016: 206) schreibt im gleichen Band wie Kecker über allgemeine Schwierigkeiten bei der Bewertung von Prüfungen, bezieht sich jedoch auf schriftliche Textproduktionen. Vorgestellt wird (2016: 208) das sog. Projekt „Bewerterkalibrierung“, welches sich jedoch ebenfalls auf schriftliche Textproduktionen bezieht. *telc* differenziert allgemein zwischen Prüfenden und Bewertenden. Beide müssen eine *telc*-Qualifizierung durchlaufen und erhalten nach der Teilnahme eine *telc*-Prüfer- bzw. Bewerterlizenz. Bewertende bewerten die schriftlichen Leistungen, Prüfende die mündlichen Leistungen. Beide müssen vor der eigentlichen Qualifizierung an dem Online-Modul „A1 bis C2 Einführung für Prüfende“ teilnehmen. Diesen Teil der Prüfung haben die Interviewten nicht absolvieren müssen, da alle bereits Prüferlizenzen für andere Niveaus besitzen, so dass dieser Teil offensichtlich neueren Datums ist. Für die vorgeschaltete Qualifizierung meldet man sich bei *telc Campus* (<https://campus.telc.net/?lang=de>) an. Angeboten wird dann ein *Moodle*-Kurs, der laut Internet in drei bis vier Stunden absolviert werden kann. Überprüft werden die Voraussetzungen, also die Inhalte des *Moodle*-Kurses, mithilfe eines Quiz. Das Quiz beinhaltet 51 Fragen. Das Quiz besteht aus sehr vielen *telc*-spezifischen Fragen (über 30 Fragen), aber auch einigen spezifischen Fragen z.B. zu Einschätzungen anhand der Skalen des *GER*. Nach erfolgreichem Abschluss des Quiz erhält man eine E-Mail:

*Vielen Dank für die Abgabe Ihrer Antworten zu ‚Vorbereitung zur Prüferqualifizierung‘ im Kurs ‚A1 bis C2 Einführung für Prüfende‘ am Mittwoch, 19. Februar 2020, 21:47. Gesamtfeedback: Sie sind sehr gut auf Ihre Prüferqualifizierung vorbereitet. Wir wünschen Ihnen einen spannenden Austausch mit den anderen Prüferinnen und Prüfern und wünschen eine erfolgreiche Teilnahme.*

Die anschließende Qualifizierung dauert acht Stunden und die Teilnehmerzahl ist auf 16 Personen beschränkt. Eine Online-Qualifizierung als Auffri-

schung ist möglich. Besitzt man bspw. eine Prüferqualifizierung für das Niveau B1–B2, die im Rahmen einer Präsenzveranstaltung erworben wurde, ist eine reine Online-Qualifizierung für eine C1 Prüfung ebenfalls möglich. Alle Qualifizierungen sind dabei kostenpflichtig. Die Prüferlizenz erhält man, wenn man alle Teilnahmevoraussetzungen erfüllt, sich aktiv am Workshop beteiligt und die mündlichen Leistungen anhand von Videoaufzeichnungen von Prüfungsteilnehmenden „angemessen“ bewertet, also der richtigen GER-Stufe zuordnet. Die Lizenz ist drei Jahre gültig. Bei der Präsenz-Qualifizierung erhält man die Handreichung Prüferqualifizierung Deutsch C1–C2. Diese beginnt zunächst mit einer Einführung in den GER und der Darstellung der Struktur der mündlichen Prüfung und den Bewertungskriterien für die mündliche Prüfung: Ausdrucksfähigkeit, Aufgabenbewältigung, formale Richtigkeit und Aussprache und Intonation. Eingegangen wird knapp auf das Prüferverhalten, und es werden sehr allgemeine Tipps formuliert. Ferner enthält die Handreichung zahlreiche Übungstests. Anhand der bereits erwähnten Videos und ausgeteilten Bewertungsbögen werden Einstufungen vorgenommen.

In den Allgemeinen Geschäftsbedingungen und der Prüfungsordnung der *telc gmbH* werden die Prüfungszentren und die Notwendigkeit eines Prüfungsverantwortlichen genannt: „Prüfungsverantwortliche müssen mit dem System der *telc* Prüfungen, einschließlich der Prüfungsordnung und der Richtlinien zur Durchführung der Prüfung, vertraut sein“ (S.10). Die Voraussetzungen für die Prüferlizenz sind Unterrichtserfahrung auf den entsprechenden GER-Stufen von mindestens 450 Unterrichtseinheiten, Vertrautheit mit den Kompetenzstufen des GER sowie mit dem handlungsorientierten, kommunikativen Ansatz für Unterricht und Leistungsmessung, Sprachkenntnisse mindestens auf Niveau C1, Kenntnisse der *telc* Prüfungsformate und der *telc* Prüfungsordnung. Eine DaF/DaZ-Qualifikation oder ein abgeschlossenes Studium ist keine Voraussetzung (vgl. Allgemeine Geschäftsbedingungen und Prüfungsordnung der *telc gmbH* S.18).

Eine Digitalisierung der mündlichen Prüfung findet nicht statt. Eine Sensibilisierung für verzerrende Effekte bei mündlichen Prüfungen erfolgt nicht.

### 1.3 DSH (*Deutsche Sprachprüfung für den Hochschulzugang*)

Auch die *DSH* ist bei der HRK und dem *FaDaF* gemäß § 3 Abs. 1 i.V. mit Abs. 6 der „Rahmenordnung über Deutsche Sprachprüfungen für das Studium an Hochschulen“ (RO-DT) registriert. Sie wird von jeder Hochschule (wobei einige Hochschulen die *DSH* abgeschafft haben) intern angeboten und ist nicht standardisiert. Es gibt zwar ein *DSH*-Handbuch und auch Vorbereitungsbücher auf die *DSH*-Prüfung, aber die Universitäten halten sich unterschiedlich streng an die Vorgaben. Schneider (2010) skizzierte bereits vor 10 Jahren die Bemühungen um die Standardisierung der *DSH*-Prüfung. Wollert & Zschill (2017) überprüfen, ob die *DSH* tatsächlich die Studierfähigkeit testet, gehen dabei aber weder gezielt auf den Bereich „Mündlichkeit“ noch auf die Prüferqualifizierung ein. 2019 wurde der Passus zur Prüferqualifikation präzisiert: Präzisierung §6(2): Aus der Formulierung zur *DSH*-Prüfungskommission wird eindeutig hervorgehen, dass alle Mitglieder der Kommission für DaF qualifiziert sein müssen. Dies war auch vor der Novellierung schon so intendiert, jedoch nicht in dieser Deutlichkeit formuliert (vgl. *FaDaF* Rundbrief 6, 2019).

Hier herangezogen wird die *DSH* der Universität Mainz. Sie gliedert sich in einen schriftlichen und einen kurzen mündlichen Teil. Das Internationale Studien- und Sprachenkolleg (ISSK) Mainz hat ein Mitglied im sog. *DSH*-Qualitätszirkel, der sich für die Einhaltung von Qualitätsstandards der *DSH* einsetzt.

Die Daten zur *DSH* an der Universität Mainz (ISSK) werden der Homepage entnommen: <https://www.studium.uni-mainz.de/dsh-und-dsh-aequivalente/> [Letzter Zugriff 19.10.2020]. Die Anforderungen werden auch festgehalten in der Ordnung für die *Deutsche Sprachprüfung für den Hochschulzugang (DSH)* an der Johannes Gutenberg-Universität Mainz vom 4. Februar 2013, zuletzt geändert am 27. Mai 2019.

Die schriftliche Prüfung dauert ca. vier Stunden und gliedert sich in vier Aufgabenbereiche: Hörverstehen, Leseverstehen und wissenschaftssprachliche Strukturen, Textproduktion sowie mündlicher Ausdruck.

Die schriftliche Prüfung umfasst die Teilprüfungen:

1. Verstehen und Verarbeiten eines Lesetextes und wissenschafts-sprachlicher Strukturen (Bearbeitungszeit: 90 Minuten einschließlich Lesezeit)
2. Verstehen und Verarbeiten eines Hörtextes (Bearbeitungszeit: 10 Minuten nach dem 1. Vortrag und 40 Minuten nach dem 2. Vortrag)
3. Vorgabenorientierte Textproduktion (Bearbeitungszeit: 70 Minuten)

Genauer betrachten wir die Angaben zur mündlichen Prüfung.

Mit der Prüfung soll die Fähigkeit gezeigt werden, studienrelevante sprachliche Handlungen (Erörtern, Bewerten, Exemplifizieren, Informieren, etc.) spontan, fließend und angemessen auszuführen und zu rezipieren sowie mit relevanten Interaktionsstrategien (Sprecherwechsel, Kooperieren, um Klärung bitten, etc.) umzugehen.

#### a) Aufgabenstellung und Durchführung

Die Dauer des Prüfungsgesprächs soll 20 Minuten nicht überschreiten. Die mündliche Prüfung besteht aus einem Kurzvortrag möglichst beschreibender Art von maximal fünf Minuten und einem Gespräch von maximal 15 Minuten. Grundlage der mündlichen Prüfung sollte ein kurzer, nicht zu komplexer und sprachlich nicht zu schwieriger Text und/oder Schaubild/Gräfik sein. Zur Vorbereitung des Kurzvortrags soll eine Vorbereitungszeit von maximal 20 Minuten gewährt werden. Gruppenprüfungen sind nicht zulässig.

#### b) Bewertung

Die Leistung ist zu bewerten nach der inhaltlichen Angemessenheit, Verständlichkeit und Selbstständigkeit der Aussagen, dem Gesprächsverhalten, der sprachlichen Korrektheit und lexikalischen Differenziertheit, der Aussprache und Intonation.

In der mündlichen Prüfung soll die Kandidatin/der Kandidat beweisen, dass sie/er imstande ist, mit Verständnis und Selbstständigkeit Vorgänge, Sachverhalte und Gedankenzusammenhänge zu erfassen, sich sprachlich damit auseinanderzusetzen sowie im Gespräch angemessen darauf zu reagieren. Die mündliche Prüfung erfolgt in Form eines Prüfungsgesprächs zu Fragen der

gewählten Studienrichtung. Dem Prüfungsgespräch kann ein entsprechender Text zugrunde gelegt werden. Ein Bewertungsraster liegt vor. Die Bewertung wird in der Prüfungsordnung in §5 geregelt; die Prüfer in §6.

Dem Prüfungsausschuss *DSH* gehören eine Vorsitzende oder ein Vorsitzender sowie zwei Stellvertreterinnen oder Stellvertreter an, die alle drei für den Bereich Deutsch als Fremdsprache qualifizierte hauptamtliche Mitarbeiterinnen oder Mitarbeiter des ISSK sind. Prüferinnen und Prüfer sind haupt- und nebenamtliche Lehrkräfte des Lehrgebiets Deutsch als Fremdsprache.

Über jede mündliche Prüfung wird eine Niederschrift angefertigt. Die Niederschrift muss die Namen der Mitglieder der Prüfungskommission, der oder des Protokollführenden und des Prüflings, Beginn und Ende der Prüfung, die wesentlichen Gegenstände der mündlichen Prüfung, die Prüfungsleistungen und das Ergebnis beinhalten. Eine Digitalisierung der Prüfung findet aktuell nicht statt, ist aber laut *FaDaF* Rundbrief vom 30.06. in Planung.

Über eine etwaige Prüferqualifizierung finden sich keine Informationen. Auch Absolventen des Masterstudiengangs DaF/DaZ in Mainz haben zwar ein Seminar zu „Prüfen und Testen“ absolviert, sind damit aber sicherlich nicht ausreichend auf die Bewertung mündlicher Leistungen vorbereitet. Der *GER*, inkl. dem *Begleitband* zum *GER* (2018), sind ihnen dagegen bekannt. Sie können also eine Einstufung nach den Stufen des *GER* vornehmen.

Es gilt demnach zu erfragen, wer die mündlichen Prüfungen abnimmt und ob diese, wie seit 2019 gefordert, DaF-qualifiziert sind.

## 1.4 Tabellarische Zusammenfassung und Entwicklung der Leitfragen

	<i>TestDaF</i>	<i>telc C1 Hochschule</i>	<i>DSH</i>
Studium	sprachlicher Hochschulabschluss (DaF, Germanistik, moderne Fremdsprache)	keine Angaben	DaF oder DaZ-Studium oder sprachlicher Hochschulabschluss
weitere Voraussetzungen	Erfahrung mit Sprachprüfungen 2–3 Jahre Lehrerbefahrung auf hoher Niveaustufe 2-tägige Schulung; inkl. Prüfer-Effekten (s.o.) regelmäßige Weiterbildungsseminare vor jeder Prüfung; Kalibrierungsunterlagen	450 Stunden Lehrerbefahrung Online-Test (Prüfer) 1-tägige Schulung regelmäßige Seminare (freiwillig) oder Lizenzen für Prüfungen auf dem B2 Niveau	bei DaF-Studium: keine sonst: 100 Stunden DaF-Qualifizierung 500 UE Lehr-/Prüfungserfahrung
Dokumentation	digitale Erfassung aller Prüfungsleistungen und der Beurteilung	Protokoll*	Protokoll
Inspektion	vorhanden	vorhanden	–

\* laut den Prüfenden unterschiedliche Angaben; vorwiegend zu Formalia: Prüfungsprotokoll A50<sup>1</sup>.

*TestDaF* fordert für die Prüfenden ein sprachbezogenes Studium; die Forderung eines DaF-spezifischen Studiums bei der *DSH* ist neuer und wahrscheinlich noch nicht flächendeckend berücksichtigt. Bei *telc* scheint ein

.....

- 1 Anmerkung einer Prüfenden: „Es gibt ein Protokoll für die gesamte Prüfung, also schriftlich sowie mündlich, ist aber nicht wirklich ein Inhaltsprotokoll, sondern dort wird eingetragen, ob alle Umschläge zu waren, wann die Prüfung anfang und zu Ende war, wie viele TN teilgenommen haben und ob es besondere Vorkommnisse gab. Außerdem werden dort nochmal die Namen der Prüfenden und telc Lizenz-Nummern angegeben. D.h., es gibt KEINE Aufzeichnungen von dem, was in der mündlichen Prüfung passiert, lediglich einmal den Bewertungsbogen, auf dem wir während der Prüfung unsere Bewertungen der einzelnen Teile eintragen können, was anschließend von den Prüfenden auf den Prüfungsbogen übertragen wird. Platz für Kommentare sind da nicht wirklich.“

sprachbezogenes Studium oder eine weitere Qualifizierung nicht obligatorisch.

Ziel der folgenden qualitativen Studie ist es nun, durch teilstrukturierte Interviews und Befragungen den Ist-Stand der Prüferqualifizierung zu erheben. Die meisten Wissenslücken in Bezug auf die Qualifizierung der Prüfenden liegen bei der *DSH* vor. Beim *TestDaF* und *telc* gilt es zu erfragen, was die genauen Inhalte der Schulung sind. Bei den nicht DaF-Absolventinnen des ISSK gilt es auch zu erfragen, ob sie die geforderten 100 Stunden DaF-Qualifizierung absolviert haben. Neben einer Befragung von VertreterInnen der Institutionen bieten Befragungen der Prüfenden einen vielversprechenden Einblick in deren Qualifizierung. Aus den o.a. Analysen wurde eine Interviewleitfaden entwickelt.

## 2 Studie

Es handelt sich hier um eine qualitative Pilotstudie, die in größerem Umfang und dann auch quantitativ durchgeführt werden müsste. Befragt wurden die Institutionen selbst, aber auch jeweils drei Prüfende. Die Ergebnisse könnten als Grundlage für eine quantitative Studie genutzt werden. Zunächst werden die Ergebnisse der Befragung der Institutionen skizziert. Es folgen die Ergebnisse der Interviews mit den Prüfenden.

### 2.1 Befragung der Institutionen

Alle drei Institutionen waren sofort zur Mithilfe bereit. Dabei stellte sich die Problematik, dass bei der *DSH* die Verantwortlichen zugleich auch die Prüfenden sind, also keine getrennte Befragung stattfinden konnte. Bei *TestDaF* und *telc* wurden die Organisation und die Prüfenden getrennt befragt. Dabei erhielten die Institutionen aus Kapitel 1 den Abschnitt zu ihrer Institution mit der Bitte um Kommentare und im Falle *telc* und *DSH* einige gezielten Fragen.

#### 2.1.1 *TestDaF*

Bezüglich der Anzahl der Testzentren wird angemerkt, dass die Zahlen Veränderungen unterliegen. Ende 2018 waren es 519 Testzentren in 99 Ländern.

Hinzugekommen sind 2018 Testzentren in Argentinien, El Salvador, Kamerun und der Russischen Föderation.

Ferner wird noch einmal präzisiert, dass die mündliche Prüfung digitalisiert wird, der Terminus „Prüfender“ also besser durch Bewertender oder Beurteilender zu ersetzen wäre. Ergänzt wird bezüglich der Auswahl der Beurteilenden, dass neben profunden Kenntnissen des GER auch explizit studien-, hochschul- und wissenschaftsbezogene Kompetenzen notwendig sind. Beurteilende werden folglich in Bezug auf diese Kompetenzen überprüft.

### 2.1.2 *telc*

Den Dokumenten war zu entnehmen, dass es keine separate Prüfung für C1 Hochschule gibt. Dies wurde von *telc* bestätigt:

*Einen gesonderten Lizenzierungskurs ausschließlich für C1 Hochschule gibt es in der Tat nicht, aber zwei Möglichkeiten eine C1 Lizenz zu erwerben, zum einen online, zum anderen über eine Präsenzveranstaltung im Rahmen einer C1–C2 Prüferqualifizierung. Wer Unterrichtserfahrung auf beiden Niveaus nachweist, erwirbt die komplette Lizenz. Alle, die nur C1 nachweisen können, erhalten die C1 Prüferlizenz.*

Aktive Teilnahme am Workshop, so *telc*, reicht nicht aus, vielmehr „müssen die Beiträge auch inhaltlich angemessen und qualifiziert sein“.

Bezüglich der fehlenden Sensibilisierung für verzerrende Elemente in mündlichen Prüfungen merkt *telc* an, dass „die Prüfertrainer/innen dazu geschult sind und das Thema bei den passenden Videos einbringen“.

Bestätigt wird, dass, anders als bei den anderen beiden Institutionen, keine DaF/DaZ Qualifizierung notwendig ist, um die Prüferqualifizierung zu absolvieren.

*Erfahrene, ältere Kursleitende haben oftmals kein relevantes Studium absolviert, sondern waren Quereinsteiger/innen. Das bedeutet weniger, dass ihre Kompetenz gering ist, sondern reflektiert die Nicht-Existenz von DaF/DaZ-Studiengängen in früheren Jahren. telc bietet allen Kursleitenden (DaF/DaZ-Absolventen/innen ebenso wie Quereinsteigenden) die*



*Möglichkeit eine Prüferlizenz zu erwerben unter der Voraussetzung, dass sie sowohl über ausreichende Unterrichtserfahrung als auch ausreichende Fachkompetenz verfügen. Bei berechtigten Zweifeln werden zusätzliche Nachweise angefordert.*

Gefragt wurde ferner nach dem Fehlen von Informationen zu den Gütekriterien in der Handreichung Prüferqualifizierung.

*Zur Vermeidung von Überfrachtung der Qualifizierung wird nicht eingehend auf dieses Thema eingegangen, ohne es jedoch gänzlich zu vernachlässigen. Reliabilität und Objektivität werden ferner im Rahmen der Herleitung der Bewertungskriterien sowie deren Anwendung bei der Bewertung von Testvideos thematisiert.*

Spezifischer wurde nach einer fehlenden Thematisierung der Verzerrung in der Handreichung gefragt.

*Obwohl in der Handreichung nicht ausführlich auf das Thema eingegangen wird, wird es dennoch indirekt im Rahmen der Diskussion um Prüferverhalten angesprochen. Viele Hinweise, besonders unter der Rubrik „Was Sie nicht tun sollten“, helfen bei der Vermeidung z. B. von Kontaminationseffekten oder schaffen Bewusstsein für den Einfluss von Persönlichkeitsfaktoren. Andere Verzerrungseffekte (z. B. den Haloeffekt) finden auf natürliche Weise Eingang in die Diskussionen über die Bewertung einzelner Videos, wenn die Seminarteilnehmenden genau in diese Falle tappen.*

### 2.1.3 DSH (ISSK)

Das ISSK merkte an, dass es sich zwar nicht um eine standardisierte Prüfung handelt, jedoch die Rahmenordnung die Standards vorgibt und ein ausführlicher Kriterienkatalog des FaDaF vorliegt. Dort sind die Standards exemplarisch dargestellt. Ferner hat das ISSK inzwischen kein Mitglied mehr im

DSH-Zirkel. Zu der Frage, wer denn prüfen darf, wurde die Prüfungsordnung zitiert, in der steht:

*Prüferinnen und Prüfer sind haupt- und nebenamtliche Lehrkräfte des Lehrgebiets Deutsch als Fremdsprache. Sie bilden eine oder erforderlichenfalls mehrere Prüfungskommissionen, die sich jeweils mindestens zur Hälfte aus hauptamtlichen Lehrkräften des Lehrgebiets Deutsch als Fremdsprache zusammensetzen.*

Allerdings wurde nicht deutlich, inwieweit eine tatsächliche DaF-Qualifizierung oder besser eine gezielte Prüferqualifizierung vorliegt.

## 2.2 Befragung der Prüfenden

Die qualitative Befragung erfolgte lediglich bei drei Prüfenden aus den jeweiligen Bereichen. Die Daten wurden teils bereinigt, ohne dabei den Inhalt zu verändern. Die einzelnen Fragen werden in der Darstellung aufgeführt. Die Ausführungen bei der ersten Interviewten sind dabei jeweils am ausführlichsten, da es bei den Personen 2 und 3 oft zu Wiederholungen kam.

### 2.2.1 *TestDaF*

**Frage 1: Was sind denn die Voraussetzungen, um Prüferin zu werden?**

*P1: Voraussetzung für eine Beurteiler\*innen-Tätigkeit ist einerseits ein einschlägiges Studium und andererseits (langjährige) Erfahrung im Bereich Deutsch als Fremdsprache (DaF).*

*P2 & P3: dito*

P1 verfügt über ein einschlägiges Studium sowie über langjährige Erfahrung mit Lehr- und Prüftätigkeit im Bereich DaF (Honorarlehrerin an TUs, Goethe-Institut, VHS; hauptamtlich im Bereich DaF im Hochschulbereich, dort auch DSH). Für das *TestDaF*-Institut ist sie seit 2002 als Beurteilerin in den Prüfungsteilen Schriftlicher Ausdruck (SA) und Mündlicher Ausdruck (MA) sowie Hörverstehen (HV) tätig und seit 2018 für die Prüfungsteile Schreiben und

Sprechen im digitalen *TestDaF*. Sie nimmt ferner an internen Kalibrierungen teil, die der Erstellung von Prüfungsaufgaben (SA, MA) sowie von Kalibrierungsunterlagen (SA, MA) für Beurteiler\*innen eines Testdurchlaufs dienen. Sie leitet zudem „testsatzübergreifende Kalibrierungen“ (TüK).

P2 hat ebenfalls ein einschlägiges Studium („Sprachlehr- und Lernforschung/Schwerpunkt DaF“) und unterrichtet in diversen Kursformaten. Sie bewertet mündliche Prüfungen seit 2017. Sie hat sowohl an der Qualifizierung als auch an der internen Kalibrierungsrunde für neue Modellsätze teilgenommen.

P3 verfügt ebenfalls über ein einschlägiges Studium und hat lange Erfahrung im Bereich des DaF-Unterrichts und hat ebenfalls an der Qualifizierung und der Kalibrierungsrunde teilgenommen.

## Frage 2: Wie sieht die Prüferqualifizierung aus?

*P1: Die Qualifizierung als Beurteilerin für den TestDaF erfolgt in den vom TestDaF-Institut durchgeführten kostenfreien Schulungen. Die bisherigen TüK, ab 2020 Zertifizierungen, finden jeweils in der Region statt, sodass eine Teilnahme ohne übermäßigen Zeit- und Reiseaufwand möglich ist. Bis einschließlich 2019 war dabei eine jeweils ganztägige Schulung für jeweils einen Prüfungsteil, also schriftlicher Ausdruck (SA), mündlicher Ausdruck (MA) und Hörverstehen (HV), vorgesehen, ab 2020 im zweijährlichen Turnus verkürzt sich die Präsenz auf einen halben Tag pro Prüfungsteil bei schriftlichem und mündlichem Ausdruck, beim HV bleibt es bei einem ganzen Tag. Jedoch minimiert sich die tatsächliche Schulung keinesfalls, da sowohl vorbereitendes Material vom TestDaF-Institut zur Verfügung gestellt wird, als auch eine nachbereitende Aufgabe oder Leistung zwecks finaler Erteilung des Zertifikats für die Beurteilung von Prüfungsleistungen im papierbasierten TestDaF erbracht werden muss.*

*P2: Die Qualifizierungen oder Schulungen sind kostenfrei und fanden bei mir an einem Wochenende statt; darüber hinaus waren „Hausaufgaben“ zu erbringen, die ein ausführliches Feedback erhielten. Ferner wurde ich nach einem ausführlichen Lebenslauf gefragt und musste eine Auflistung*

*der unterrichteten Unterrichtseinheiten auf den Niveaustufen B2 bis C2 vorlegen. Jährlich finden Nachschulungen statt.*

*P3: Die Beurteilerschulung findet jährlich statt, die Auffrischungen ebenfalls regelmäßig.*

**Frage 3: Was sind die Inhalte der Qualifizierung, wie die empfundene Qualität der Qualifizierung und fühlen Sie sich gut vorbereitet?**

*P1: Die Inhalte der Qualifizierung zielen darauf ab, ein größtmögliches Verständnis für die Beurteilungsvorgaben des TestDaF-Instituts zu erreichen und das Beurteilungsverhalten der Beurteiler\*innen dahingehend zu optimieren. Zu diesem Zweck werden Schwierigkeiten hinsichtlich der Beurteilung einzelner Aufgaben des SA, MA, HV und einzelner Beurteilungskriterien thematisiert und diskutiert. Grundlage dafür sind Prüfungsleistungen aus früheren Testläufen. Die Rückmeldung aus und der rege Austausch mit dem TestDaF-Institut, in der Präsenzschulung vertreten durch die Schulungsleiter\*innen, sowie der Erfahrungsaustausch mit anderen Beurteiler\*innen sind ein wichtiger Bestandteil der Zertifizierungen. Hierdurch fühlt man sich als Beurteiler\*in alljährlich/zweijährlich „auf Kurs gebracht“, eben kalibriert. Dies vermittelt Sicherheit für die kommenden Beurteilungssituationen. Auch die Rückmeldungen bezüglich Konsistenz und Strenge der eigenen Beurteilungen nach einem jeden Testlauf sind ein wichtiger Faktor, der Beurteiler\*innen die für die Arbeit notwendige Sicherheit gibt. Sollte man sich in einer Beurteilungsphase trotz allem mit Problemen konfrontiert sehen, ist eine Kontaktaufnahme mit den Referatsleiter\*innen der Weg, der allen Beurteiler\*innen offensteht. Dies ist ein weiterer Faktor, der Sicherheit vermittelt. Wie zu sehen, fühle ich mich ausreichend geschult und gut vorbereitet auf die Arbeit als Beurteilerin.*

*P2: Die Qualifizierung ist jeweils nach schriftlich und mündlich getrennt. Es erfolgt eine Vorstellung des Testformats und des Aufbaus, im Detail: Teile der Aufgaben, Erwartungshorizont (Was sollen die Prüfungsteilnehmenden leisten?), Bewertungskriterien, einzelne Abstufungen, Honorarabrechnung, statistische Auswertung und den internen Ablauf. Man hatte*

*noch immer großen Respekt vor der Bewertung und Durchführung, aber durch das individuelle Feedback, die Erklärung der Statistik, dass man da nicht allein verantwortlich ist, hat sich das abgemildert. Man bekommt nach jedem Bewertungsdurchgang ein Feedback. Ich fühle mich also sehr gut vorbereitet.*

P3: Die inhaltliche Darstellung ist identisch mit den beiden o.a. Berichten. Daher erfolgt hier die transkribierte Antwort auf die Frage: „Fühlten Sie sich denn gut vorbereitet?“

*Auf alle Fälle. Wichtig ist jedoch neben der Schulung und den Auffrischungen, dass man auch wirklich bewertet und sein Wissen anwendet. Die Kombination Schulung, Auffrischung und Erfahrung und der jährliche Austausch sind ideal. Man denkt immer vor der Auffrischung: Muss das wirklich sein? Aber ich bin immer wieder froh, weil sich manchmal Dinge auch verschieben oder der Blick auf manche Kriterien ist zu starr geworden und der Austausch führt dazu, dass man manches überdenkt. Also die Qualifizierung ist auf jeden Fall zu empfehlen. Ich fühle mich sehr gut betreut.*

**Frage 4: Was hat Ihnen bei der Prüferqualifizierung am meisten geholfen?**

*P1: Es ist nicht ein Faktor auszumachen, der am meisten hilft, es ist das Gesamtpaket an Maßnahmen, die das TestDaF-Institut trifft und bereitstellt.*

*P2: Die Vorbereitung auf das Material, das individuelle Feedback, die genaue Erläuterung der Aufgaben und des Erwartungshorizonts.*

*P3: Die gesamte Qualifizierung mit dem ausführlichen Feedback und dem regen Austausch.*

**Frage 5: Wie sind Sie zu einer guten Prüferin geworden?**

*P1: Intensive Vorbereitung und Betreuung durch das TestDaF-Institut waren und sind ein wichtiger Teil der Qualifizierung. Jedoch helfen die*

*besten Maßnahmen nicht, wenn nicht auch die Bereitschaft besteht, sich selbst in seinem Beurteilungsverhalten immer wieder in Frage und auf den Prüfstand zu stellen. Ich verfüge NICHT über eine Selbstzufriedenheit als Prüferin. Das hilft mir, offen zu bleiben für Verbesserung und Optimierung.*

*P2: Der Austausch mit Kollegen/-innen durch Feedback und Nachschulungen sowie die Bearbeitung der Modellprüfungen.*

*P3: Wie bereits dargelegt: die Kombination aus Erstqualifizierung, dem eigentlichen Bewerten, den Auffrischungen und dem Austausch mit den Kollegen und Kolleginnen.*

**Frage 6: Wird die Prüfung aufgezeichnet? Wenn nicht, wird ein ausführliches Protokoll angefertigt? Was wird im Protokoll vermerkt?**

*P1: Zu dieser Frage müsste die prüfende Institution befragt werden. Hinsichtlich der Beurteilung von Prüfungsleistungen ist zu sagen, dass das TestDaF-Institut von den Beurteiler\*innen ausreichende Begründungen dafür erwartet, mit welcher Niveaustufe (TDN) einzelne Items auf Grundlage der Kalibrierungsunterlagen zum jeweiligen Testlauf beurteilt werden. Auf diese Weise wird die Beurteilungsarbeit der einzelnen Beurteiler\*innen dokumentiert.*

*P2: Die mündliche Prüfung wird aufgezeichnet, die Bewertung findet zu Hause statt und man füllt einen Bewertungsbogen mit präzisen Beispielen aus.*

*P3: dito*

**Frage 7: Wie viele Bewertende bewerten eine Prüfung?**

*P1: Auch zu dieser Frage müsste die prüfende Institution befragt werden.*

*P2: Eine Bewertung, aber mit statistischer Bereinigung; bei Zweifelsfällen kommt ein zweiter Bewerter hinzu.*

*P3: dito*

**Frage 8: Nach welchen Kriterien beurteilen Sie denn mündliche Texte? Wie wurden Sie in der Qualifizierung diesbezüglich geschult?**

*P1: Das habe ich ja bereits bei den Inhalten der Qualifizierung und der Frage nach dem gut vorbereitet sein beantwortet. Die Kriterien liegen vor, werden intensiv behandelt und sind somit zentraler Bestandteil der Qualifizierung.*

*P2: Nach den gegebenen Bewertungskriterien und immer vor dem Hintergrund: Kann diese Person mit diesem aktuellen Sprachstand ein Studium in D aufnehmen und im „Unidschungel“ bestehen und sich weiterentwickeln. Dies wurde in der Qualifizierung ausführlich geschult und es gab immer wieder Rückmeldung zur Bewertung.*

*P3: Also wir haben ja unsere Kriterien, wo wir jede mündliche Aufgabe nach diesen acht Kriterien bewerten und meistens geht es dann darum, diese zu hinterfragen, also genauer zu überprüfen, wie wurde das umgesetzt, wie kann ich das für mich noch ein wenig eingrenzen und an konkreten Beispielen gucken, wie habe ich mir vielleicht diesen Deskriptor so schön zurechtgelegt, der dann doch nicht so ganz passt. Wichtig ist, sich immer zu hinterfragen.*

**Frage 9 & 10: Was versteht man denn unter dem sog. Positionseffekt? War das Bestandteil der Prüferqualifizierung? Und was verbirgt sich hinter dem Halo-Effekt? Wurde dies in der Prüferqualifizierung behandelt?**

*P1: Der Positionseffekt und auch der Halo-Effekt sind Phänomene, die in den Zertifizierungen immer wieder angesprochen werden, da diese Effekte nur allzu menschlich sind, aber selbstverständlich weitestgehend vermieden, zumindest aber minimiert werden müssen. Denn nur so kann eine faire, allen Prüfungsteilnehmer\*innen gerecht werdende Beurteilung ihrer Leistungen gewährleistet werden. Bei einer Beurteilung von Prüfungsleistungen in einem Umfang von 20, 50 oder gar mehr Leistungen ist der Positionseffekt eine stetige „Gefahr“. Haben wir beispielsweise eine Reihe sehr schlechter Leistungen beurteilen müssen, freuen wir uns über die eine Leistung, die besser ist, und neigen dazu, diese zu gut zu beurtei-*

*len. Vermieden werden kann dies am ehesten, wenn jede Prüfungsleistung einzeln betrachtet wird und keinerlei Rückblicke auf und Vergleiche mit vorherigen vorgenommen werden.*

*Auch der Halo-Effekt muss unbedingt vermieden werden. Die Anonymisierung von Prüfungsleistungen mittels TN-Nummern ist solch ein Versuch der Vermeidung. Jedoch verraten Schriftbild (beim SA oder HV) und/oder Akzent (beim MA) erfahrenen Beurteiler\*innen oft recht deutlich, welche Nationalität, Ausgangssprache und typische Schwierigkeiten und Fehler die Prüfungsteilnehmer\*innen zeigen werden. Eine mögliche Vor-Beurteilung, der Halo-Effekt, steht also im Raum. Vermieden werden kann dies nur dadurch, dass man sich dieser typischen Schwierigkeiten und Fehler vielleicht bewusst ist, diese aber nicht als Rechtfertigung für gute und/oder schlechte Leistungen heranzieht, beispielsweise „für einen XY ist das aber schon ganz gut“ oder „das können XY halt nicht besser“. Im Fokus der Beurteilung von Sprachleistungen sollte eher die Verständlichkeit und Korrektheit der Performanz stehen.*

*P2: In der Schulung ist man sehr intensiv auf diese Effekte eingegangen und wir haben immer wieder über das Prüferverhalten gesprochen und dass man Dinge auch mal „liegen lassen“ und neu starten muss.*

*P3: Auf jeden Fall. Auch schon in der Erstqualifizierung wird da sehr viel Wert draufgelegt.*

### **Frage 11: Was möchten Sie uns eventuell noch über die Qualifizierung berichten?**

*P1: Nichts, es scheint mir vieles gesagt. Bei weiteren Nachfragen stehe ich aber gern zur Verfügung.*

*P2: Das digitale Format ist in Bezug auf das Aufgabenformat sehr viel näher an den Uniaufgaben, die Studierende zu bewältigen haben. Es wird viele Teilnehmende vor Herausforderungen stellen, weil sie es nicht gewohnt sind, aber es zeichnet ein realistischeres Bild. Viel zu viele Zertifikate, egal welches, werden zu leichtfertig vergeben und ich habe die Studis dann in meinen Deutschkursen an der Uni sitzen und darf noch einmal eine Satzklammer etc. erklären. Gleiches gilt aber auch für normale Kur-*



*se. Lehrkräfte sollten öfter mal ‚nein‘ zu einer Weiterstufung sagen, auch wenn es den Teilnehmenden nicht gefällt.*

*P3: Ich kann nur sagen, dass ich das [die Qualifizierungen] immer als extrem gut organisiert und vorbereitet finde und mich da richtig gut aufgehoben fühle, also im Ganzen. Da ist einfach eine gute Atmosphäre, um da zu arbeiten und auch von der Konstellation passt es immer. Manchmal trifft man die gleichen Beurteiler, mit denen man sich sehr gut austauschen kann, manchmal neue – immer anregend. Es ist auf jeden Fall eine sehr fruchtbare Zusammenarbeit.*

### 2.2.2 *Telc*

#### **Frage 1: Was sind die Voraussetzungen, um Prüferin zu werden?**

*P1: Wenn ich das richtig in Erinnerung habe, muss man lediglich Unterrichtserfahrung von 450 Stunden und Kenntnisse mindestens auf dem gleichen Niveau haben, das man dann prüft und natürlich die Qualifizierung absolvieren.*

*P2: Unterrichtserfahrung auf dem Niveau C1. Aber ich wurde auch nicht gefragt.*

*P3: Danach wurde ich nicht gefragt, aber ich habe ja ein DaF-Studium, braucht man aber glaub nicht unbedingt.*

P1 hat selbst sowohl einen Fernstudienkurs DaF am Goethe-Institut/Universität München absolviert, ein abgeschlossenes Masterstudium DaF/DaZ und arbeitet seit 2014 im Bereich DaF auf allen Niveaustufen und das sowohl im universitären Bereich als auch im Integrationsbereich. *telc*-Prüferqualifizierungen (DTZ, B1–B2 und C1–C2) hat sie 2017 und 2018 erworben.

P2 hat ebenfalls ein einschlägiges Studium. Die *telc*-Prüferlizenz hat sie seit 2012 und das auf allen Niveaustufen (B1/B2, C1/C2 und C1+ Hochschule sowie Briefbewertung).

P3 hat ebenfalls DaF studiert und die Prüferlizenz seit 2014, ebenfalls auf allen Niveaustufen.

**Frage 2a: Was sind die Inhalte der Qualifizierung, wie die empfundene Qualität der Qualifizierung [und fühlen Sie sich gut vorbereitet? – hier als separate Frage]**

*P1: Die Erstschulung umfasst 8 Stunden und die Inhalte der Qualifizierung entsprechen den offiziellen Angaben auf der Homepage; also eine Vertiefung des Wissens über den GER, das Kennenlernen der Kriterien für die Bewertung der mündlichen Leistungen sowie das Diskutieren und Bewerten mündlicher Leistungen anhand von Videoaufzeichnungen; dabei wird auch das Prüferverhalten analysiert. Daneben gibt es eine online-Qualifizierung. Die Schulungsreferent\*innen waren gut vorbereitet und erfahren, sowohl als Referenten als auch selbst als Prüfer\*innen. Die Teilnehmergruppe bei der Erstschulung schien mir, erstaunlicherweise, als heterogen bezüglich wichtigem Vor- und Grundlagenwissen, z.B. hinsichtlich des Prüfens allgemein. Daher war es schwierig, alle „richtig abzuholen“, und somit erscheint mir die Schulung an nur einem Tag zu wenig. Die Auswahl der für das Bestehen zu bewertenden Videos erschien mir manchmal zu „einfach“, solche „Paradebeispiele“ an (guten) Prüflingen hat man in der Realität eher selten. Hier wäre wirklich authentischeres Material hilfreich. Problem dabei: Die Prüflinge müssen der Videoaufnahme zustimmen, schlechtere TN machen das weniger gern oder nicht. Ich bezweifle auch, dass man mit den drei Lizenzen online eine Lizenz im Bereich „Pflege“ machen kann. Hier sollte ein längerer Präsenz-Termin obligatorisch sein.*

*P2: Die Erstschulung ist acht Stunden und sie ist leider kostenpflichtig. Eine Nachschulung habe ich nach drei Jahren absolviert, aber auch die ist kostenpflichtig (4 Stunden à 40€). Die Inhalte umfassten die GER-Skalen, den Prüfungsaufbau, die Bewertungskriterien, einige Prüfungsbeispiele mit dem Vergleich der eigenen Einschätzung mit der telc-Einschätzung. Die eigene Prüfungsbewertung hat der Dozierende mit nach Hause genommen und später erfuhr man, ob man die Lizenz bekommt oder nicht. Das machte einen ziemlich willkürlichen Eindruck.*

*P3: Wie auch auf der Homepage angegeben, acht Stunden. Der Fokus liegt auf den GER-Skalen, dem Prüfungsaufbau und den Bewertungskriterien.*

*Mittels Videos von Prüfungen wird diskutiert. Leider fehlt ein Feedback zu der eigenen Einschätzung.*

**Frage 2b: Fühlen Sie sich gut vorbereitet?**

*P1: Gut vorbereitet fühlte ich mich vor allem nach der 1. Schulung nicht. Nach der Theorie und den Videos weiß man einfach noch nicht, was es heißt, eine Prüfung selbst abzunehmen. Aber man ist ja zu zweit und kann v.a. bei den ersten Prüfungen den „passiven“ Teil übernehmen (= „nur“ bewerten, nicht die Prüfung leiten). Mit gesammelter Prüfungserfahrung kann man in den Schulungen gut diskutieren oder Fragen klären. Daher ist es auch gut, trotz Möglichkeit einer Online-Schulung, in regelmäßigen Abständen zu Präsenz-Schulungen zu gehen.*

*P2: Ich fühlte mich nach der Prüferqualifizierung sehr in der Luft hängend, vor allem da man auf die finale individuelle Prüfungsbewertung nie ein Feedback bekommt. Auch heute bekommt man keine Rückmeldung über Strengegrad, Durchführung, etc. Veränderte Regularien bekommen auch immer nur die Testzentren, Prüfende haben Glück, wenn sie den Newsletter abonniert haben und es da drinsteht.*

*P3: Nein, die Informationen im Studium dazu waren ausführlicher. Mit dem Wissen aus dem Studium, den Schulungen und dem Austausch mit dem Zweitprüfenden erlangt man aber nach einer gewissen Zeit eine gewisse Prüfungssicherheit.*

**Frage 3: Was hat Ihnen bei der Prüferqualifizierung am meisten geholfen?**

*P1: Beispielvideos von realen Prüfungen anzuschauen und die Bewertung im Plenum zu besprechen, und vor allem bei Uneinigkeit zu diskutieren. Der Austausch mit anderen TN/KL zu Schwierigkeiten beim Prüfen (v.a. bei der allerersten Schulung, wenn man selbst noch keine Prüferfahrung hat) und das konkrete Ausfüllen der Original-Bewertungsbögen, wie in der Prüfung selbst.*

*P2: der Austausch mit Kolleginnen.*

*P3: Wie eben schon gesagt, die Kombination aus Studium, der Schulungen und dem Austausch mit den anderen Prüferinnen.*

**Frage 3: Wie sind Sie zu einer guten Prüferin geworden?**

*P1: „learning by doing“, also durch Erfahrung und Durchführen von Prüfungen, Austausch mit Prüferkolleg/innen, also eine Kombination aus Wissen, Einhaltung der Richtlinien und „Menschenverstand“.*

*P2: Der Austausch mit den Kolleginnen, das Bewerten während der Prüfung mit Kolleg\*in und „richtige“ Qualifizierungen zum mündlichen Prüfen.*

*P3: Übung macht den Meister – ich denke, dass ich inzwischen recht sicher beim Prüfen bin.*

**Frage 4: Wird die Prüfung aufgezeichnet? Wenn nicht, wird ein ausführliches Protokoll angefertigt? Was wird im Protokoll vermerkt?**

*P1: Nein, keine Aufzeichnung (weder Ton noch Video), aber ausführliches Protokoll zu mündlichen Leistungen der Prüflinge zu jeder Aufgabe und anschließender Absprache und Vergleich mit Co-Prüfer\*in für die Endbewertung. Je nach Prüfung müssen sich die Prüfenden bei der Bewertung nicht einig sein, man nimmt dann den Mittelwert.*

*P2: Nein, es gibt ein Protokoll, in dem die Prüfungscoordination Unregelmäßigkeiten notiert werden. „Gelbe“ Prüferbewertungen bleiben bei der Sprachschule, aber was ein Prüfer dort aufschreibt, ist ihm/ihr überlassen. Je nach Prüfenden also entweder ein ausführliches Protokoll oder aber sozusagen nichts.*

*P3: Jein, es gibt ein Protokoll, aber je nach Prüfendenkonstellation wird mehr oder weniger notiert.*

**Frage 5: Wie viele Bewertende bewerten eine Prüfung?**

*P1: Immer zwei: davon mindestens ein/e Externe/r, d.h. ein Prüfender, der nicht an der Institution arbeitet, an der die Prüfung abgenommen wird.*

*Oder aber, wenn er die Prüflinge seit mindestens 6 Monaten nicht mehr unterrichtet, damit Objektivität gewährleistet ist.*

*P 2: Zwei Bewertende.*

*P3: dito*

**Frage 6: Nach welchen Kriterien beurteilen Sie denn mündliche Texte? Wie wurden Sie in der Qualifizierung diesbezüglich geschult?**

*P1: Nach den Kann-Beschreibungen des GER sowie den Bewertungskriterien und -kategorien von telc. Alle Anforderungen sind schriftlich formuliert, in den Handbüchern, die man in den Schulungen ausgeteilt bekommt. Diese können auch während der Prüfung nachgelesen werden. In der Schulung selbst findet eine ausführliche Vorstellung, Bearbeitung und Besprechung des GER statt, von Grobskala bis zu einzelnen Kann-Beschreibungen der Fertigkeiten. Der Fokus liegt natürlich auf mündlicher Produktion (in der Schulung: z.B. Zuordnung der Kann-Beschreibung auf Karten nach Niveau, gemeinsames Festhalten der wichtigsten Unterscheidungsmerkmale der Niveaustufen, also z.B. Was ist gutes B1, aber noch nicht B2?).*

*Beim Ansehen der Videos und dem gemeinsamen Bewerten der Beispielprüfungen werden die telc-Bewertungen besprochen. Dabei kommt es jedoch häufig zu einer „Nichtübereinstimmung“ der Teilnehmenden mit den telc-Referenten und es wird diskutiert. Dabei bewerten mal die Teilnehmer, mal telc „zu nett“. Die Teilnehmenden in der Qualifizierung vermischen oft sprachliche und inhaltliche Ebene und sanktionieren doppelt. Ich sehe hier folgende Schwierigkeit: die telc-Musterbewertung richtet sich zwar nach dem GER, wird aber im Einzelnen von einem relativ kleinen Kreis von telc-Prüfern und Schulungsreferent\*innen vollzogen; das Team wird laut telc auch regelmäßig geschult und sie müssen unabhängig voneinander immer wieder Prüfungen bewerten, um zu kalibrieren, aber eine gewisse Heterogenität ist schon festzustellen.*

*P2: Nach den vorliegenden Bewertungskriterien. Die Schulung zum Bewerten an sich wäre durchaus ausbaufähig. Einigen Teilnehmenden würde ich, auch wenn sicher dumm für sie, die Lizenz nicht ausstellen.*

*P3: Die Bewertungskriterien wurden sehr ausführlich diskutiert und grundsätzlich würde ich die Schulung hier als sehr hilfreich ansehen. Aber hier wird die Heterogenität der Teilnehmenden sehr deutlich. Bei einigen hat man den Eindruck, es wäre ihr Erstkontakt mit dem GER oder dem Hochschulbereich.*

**Frage 7: Was versteht man denn unter dem sog. Positionseffekt? War das Bestandteil der Prüferqualifizierung? Und was verbirgt sich hinter dem Halo-Effekt? Wurde dies behandelt?**

*P1: Keine Ahnung – die Auswirkung der Anordnung der Fragen auf die Beantwortung in der Prüfung? Das war nicht Bestandteil der Schulung, falls doch wurde das so nicht genannt. Oder „Position beziehen“ in der Rolle als Prüfer/in? Also, das Prüferverhalten hat Auswirkung auf Beantwortung seitens der TN? Das wurde in der Schulung nicht thematisiert. In der Schulung wird bzgl. skaliertes Prüfungen besprochen, dass man zunächst mit Fragen des niedrigeren Niveaus beginnen soll und ggf. (je nach Niveau des TN) dann Fragen stellen, die das höhere Niveau testen (z.B. bei DTZ sind diese vorgegeben). Außerdem werden auch Dos and Don'ts eines Prüfers besprochen und wie man sich als Prüfer\*in verhalten sollte. Hm, Halo-Effekt, vielleicht Fehler bei der Personen-/Leistungsbeurteilung durch dominantes Merkmal, womit anderes in den Hintergrund gerät bzw. nicht mehr berücksichtigt wird und letztlich zur „falschen“ Beurteilung führt? Bei keiner meiner Schulungen wurde dieser Begriff explizit thematisiert, auch wenn wir uns mit dem allgemeinen Prüferverhalten beschäftigt haben. Wobei ... bei der ersten Schulung in einer Beispielprüfung im Video ein Prüfling eine sehr „komische“ Stimme hatte oder so, jedenfalls ein physisches Merkmal, das einen auf Dauer irritierte. In der Diskussion stellte sich heraus, dass die meisten der Schulungs-TN sich davon beeinflussen ließen und eine „schlechtere“ Bewertung gaben. Wobei die sprachliche und inhaltliche Leistung nicht darunter litt (das Video wurde nochmals gezeigt).*

*Außerdem, wie oben erwähnt, wird in den Schulungen, anhand der Beispiele darauf hingewiesen, dass man die inhaltliche und sprachliche Ebe-*

*ne nicht vermischen solle. Also „nur“, weil jemand ständig den gleichen Grammatikfehler macht, kann die Aufgabe dennoch inhaltlich gut erfüllt worden sein. Also nur Abzug bei „sprachlicher Korrektheit“, nicht aber bei „Aufgabenerfüllung“.*

*P2: (zu beidem): Wenn bei einer meiner Qualifizierungen jemand danach gefragt hat, wurden die Effekte angesprochen, jedoch je nach Trainer/in sehr unterschiedlich intensiv behandelt. In der eigentlichen Qualifizierung werden die Effekte nicht behandelt.*

*P3: Kann ich nicht wirklich etwas zu sagen. Sehr ausführlich behandelt wurde, dass wir Sprache und Inhalt getrennt bewerten müssen, also Inhalt kann auch gut sein, wenn die Grammatik daneben ist.*

#### **Frage 8: Was möchten Sie uns eventuell noch über die Qualifizierung berichten?**

*P1: Es gibt klare Voraussetzungen, um an einer Prüferqualifizierung teilnehmen zu können und somit lizenzierte/r Prüfer/in zu werden. Es werden bei der Anmeldung für die Schulungen jedoch keinerlei Nachweise dafür verlangt. Man muss nur einmalig (ich glaube zur ersten Präsenzschulung) vorab einen „Test“ online ausfüllen. Dieser beinhaltet Fragen zu den telc-Richtlinien und dem GER. Die Informationen dazu werden als pdf bereitgestellt und können zum Ausfüllen natürlich benutzt werden und können oft selbsterklärend beantwortet werden (bestimmt auch, wenn man nicht unbedingt vom Fach ist). Dann braucht man sich in der Schulung auch nicht zu wundern, welche Leute dort so sitzen (wirken unerfahren und haben z.B. vom GER noch nichts gehört bzw. nur aus dem Vorbereitungsteil von telc selbst).*

*Ich finde die Lizenzerteilung nicht wirklich professionell und transparent. Als Voraussetzung gilt die Teilnahme der Präsenzschulung, das richtige Bewerten einer Prüfung (per Video) und „aktive Mitarbeit“ bei der Schulung selbst. Man kann an der Schulung teilnehmen, auch ohne Nachweise der geforderten Voraussetzungen. Die Bewertung der Prüfung findet nach den gemeinsam besprochenen Prüfungen statt. Das soll zwar in Einzelarbeit passieren, aber das Vergleichen mit dem Nachbarn und das*

*Ausbessern ist möglich und gängig (wird vom anwesenden Referenten nicht wirklich unterbunden, ggf. auch von Referenten abhängig). Die für die Lizenz zu bewertende Prüfung ist meist eine sehr „einfache“, also vom Niveau her klar zuordnungsbar Prüfung; eine Abweichung von der Musterlösung innerhalb der gleichen Niveaustufe ist geduldet, doch auch bei nichtzutreffender Bewertung ist die Erteilung der Lizenz möglich, da die „aktive Mitarbeit“ (Fragen beantworten, an Diskussion teilnehmen...) anscheinend stärker gewichtet wird. Das bekommt man als TN aber eigentlich gar nicht mit, der Referent muss zu jeder/jedem TN einen Bogen ausfüllen, was m.E. bei der großen Anzahl der TN oft schwer möglich ist, während des Präsentierens selbst. Darüber werden die TN zwar zu Beginn informiert und auch zur aktiven Teilnahme aufgefordert, dennoch weiß man als TN am Ende nicht wirklich, worin man nun „gut“ war. Man glaubt automatisch, man habe die mündliche Bewertung richtig gemacht. Wie oben beschrieben findet eine Prüfung immer zu zweit statt. Man hat also die Möglichkeit einige Prüfer-Kolleg\*innen kennenzulernen. Dabei wird deutlich, dass je nach Verhalten und „Einstellung“ der Prüfer\*innen, die Prüfung ganz unterschiedlich und auch nicht immer „standardisiert“ verlaufen kann, also z.B. die Frageweise, bei ständigen Ja/Nein-Fragen einer Prüferin bekommt man wenig zu bewerten. Wobei man ja meinen sollte, wenn alle gleich geschult wurden, sollten auch die Prüfungen „gleich“ ablaufen...*

*P2: Meiner Meinung nach sollten Qualifizierungen kostenlos sein, denn man führt einen Service für diese Institution durch und wenn man schon bezahlt, sollten Kenntnisse geprüft und Feedbacks gegeben werden, so dass man sich besser aufgehoben fühlt. Bei telc erneuere ich nur noch Lizenzen, bei denen ich weiß, dass ich bei einer Sprachschule prüfen oder bewerten kann. Die Betreuung der Prüfer/Bewerter durch telc selbst ist extrem schlecht. Das Honorar wurde zwar angehoben, aber dennoch machen nur die wenigsten Lehrkräfte eine Bewerter\*innen-Schulung, da es sich nicht lohnt und auch die negativen Erfahrungen dominieren. Welche andere Institution setzt bitte Aufgaben aus Modellsätzen in realen Prüfungen ein?*

*P3: Mir haben die Schulungen Spaß gemacht und ich habe auch immer etwas Neues mit nach Hause genommen. Es waren immer auch spannende*



*Leute dabei. Etwas nervig ist es, wenn zu viele der Teilnehmerinnen recht wenig Ahnung von DaF haben, also Lernvariablen, Lerntraditionen und vor allem Methoden.*

### 2.2.3 DSH am ISSK

#### Frage 1: Was sind die Voraussetzungen, um Prüferin zu werden?

*P1: Voraussetzung ist ein Studium im Bereich Germanistik, am besten in Ergänzung mit einem DaF-Abschluss.*

*P2: Studium und Anstellung am ISSK.*

*P3: Studium und Anstellung am ISSK, also keine Lehrbeauftragten.*

P1 hat sowohl Germanistik als auch DaF studiert und unterrichtet seit 2001 DaF. Im Studium selbst wurde kein Modul zum Testen und Prüfen angeboten, so dass sie sich diesen Bereich autodidaktisch, durch Austausch mit Kollegen und Kolleginnen, Weiterbildungen und Tagungen erschlossen hat.

P2 hat ebenfalls ein DaF-Studium (Master) absolviert, prüft aber erst seit ca. 1 Jahr. Es gab in ihrem Studium kein komplettes Seminar zum Prüfen und Testen, jedoch verschiedene Sitzungen zu Testformaten. In einem Seminar zu Unterrichtskonzepten wurde jedoch darüber hinaus spezifischer auf die DSH und die Prüfungsaufgaben eingegangen, ohne dies jedoch explizit als DSH-Prüfungsformate zu benennen.

P3 hat Germanistik mit Schwerpunkt Literaturwissenschaft studiert und war als DAAD Lektor in Polen und China im Bereich DaF tätig (seit 2005), der erste DaF-Kurs fand in China statt und hier hat P3 auch Vorbereitungskurse für *TestDaF* gegeben. Seit 2017 ist P3 am ISSK festangestellt. Eine Prüferqualifizierung, die jedoch als wenig zielführend beurteilt wird, fand bei *telc* statt.

#### Frage 2: Haben Sie eine externe Prüferqualifizierung gemacht?

*P1: nein.*

*P2: nein.*

*P3: ja, bei telc (2015), aber nur A1 (und der „Workshop“ war nicht wirklich zielführend)*

**Frage 3: Wie sind Sie zu einer guten Prüferin/einem guten Prüfer geworden?**

*P1: Da ich selbst die vorbereitenden DaF-Kurse anbiete, die schließlich zur DSH führen (also die sog. „studienvorbereitenden Kurse“), und ich somit die genauen Ziele und Erwartungen in Anlehnung an den GER kenne, kann ich dieses Wissen sehr gut auch bei der Prüfung berücksichtigen. Ich kann so sehr gut beurteilen, ob die Ziele, also das Niveau auch nach dem GER, erreicht wurden. Vorteilhaft dabei ist, dass ich auch die Entwicklung der Prüflinge beurteilen kann. Mit ist jedoch bewusst, dass dies die Bewertung beeinflussen könnte. Das versuche ich zu verhindern.*

*P2: Ich bin noch auf dem Weg dahin, die vielen Diskussionen mit der Prüferin helfen dabei aber sehr und ich unterrichte ja auch am ISSK.*

*P3: Ich habe bereits viel Erfahrungen mit mündlichen Prüfungen im Ausland gesammelt und seit 2017 dann auch mit der DSH am ISSK, wobei ich zunächst Protokollant war. Beim ersten Mal, wie immer, ist man beim Protokollführen noch nicht so geübt. Ich habe gelernt, wie wichtig es ist, wirklich alles sehr genau und konsequent zu protokollieren. Ich habe mir viel von den Kollegen abgeschaut und man braucht auch die Erfahrung und man darf die mündliche Prüfung nicht als Anhängsel sehen. Der Austausch mit den Kolleginnen ist immer ein Vorteil.*

**Frage 4: Wird die Prüfung aufgezeichnet? Wenn nicht, wird ein ausführliches Protokoll angefertigt?**

*P1: Nein, aufgezeichnet wird nicht, aber gerade bei der mündlichen Prüfung wird ein sehr ausführliches Protokoll angefertigt, das dann bei der Bewertung herangezogen werden kann. Formalia werden ebenfalls im Protokoll festgehalten, das „Verlaufsprotokoll“ ist aber sehr viel relevanter und nachvollziehbar.*

*P2: Nein, aber ich mache mir sehr viele Notizen und aufgrund der Notizen und dem Gesamteindruck wird dann diskutiert. Dabei werden die Gründe für die Punktevergabe nicht zwingend notiert.*

*P3: Wie oben schon beschrieben, ist ein ausführliches Protokoll sehr wichtig, um auch den Studierenden sagen zu können, warum sie so bewertet wurden.*

**Frage 5: Wie viele Bewertende bewerten eine Prüfung?**

*P1: Jeweils zwei Personen, einmal der/die Prüfer/in und ein/eine Protokollantin, der/die aber auch immer selbst Prüfer ist. Die Bewertung erfolgt durch beide und das unter Zuhilfenahme des Protokolls. Wir diskutieren hier sehr ausführlich.*

*P2: dito*

*P3: dito*

**Frage 6: Nach welchen Kriterien beurteilen Sie denn mündliche Texte? [Wie wurden Sie in der Qualifizierung diesbezüglich geschult?]**

*P1: Wie dargestellt, gibt es keine spezifische Prüferqualifizierung. Die Kriterien sind jedoch standardisiert in einem Kriterienkatalog festgehalten. Es können 100 Punkte vergeben werden, 100 Punkte wären dann 100%.*

*P2: Wir haben einen sehr ausführlichen Bewertungsbogen und differenzieren dann sehr genau nach Inhalt und Sprache; auch Aussprache – aber nur am Rande – wird bewertet.*

*P3: Wir haben unseren Bewertungsbogen, wir sprechen uns sehr ausführlich über die Bewertung ab und wir haben auch über eine virtuelle mündliche Prüfung nachgedacht. Das hätte aber logistische Hürden zum einen und wir hätten einen anderen Bewertungsbogen entwickeln müssen. Im Vergleich empfinde ich, zumindest unsere DSH, als das beste Prüfungsformat. Die Anerkennung von telc Hochschule machen wir „zähneknirschend“. Ich kenne alle drei Prüfungen und bin überzeugter Vertreter der DSH, denn ich denke, es ist die Prüfung, die die Studierfähigkeit am exaktesten nachprüfen kann (bezogen auf die gesamte Prüfung, nicht das Mündliche), weil sie in der Aufgabenstellung zwischen geschlossen, halb-geschlossen, halb-offenen und offenen Fragestellungen am besten die Kompetenzen abbildet. Man überprüft z.B. nicht nur das reine Textverständnis z.B. beim Leseverstehen, sondern auch, ob sie darüber reflektieren können, also nicht reine Informationsentnahme, sondern auch die Frage, wie ihre Haltung dazu ist. Können die Prüfungskandidaten auch zwischen den Zeilen lesen?*

### Frage 7: Berücksichtigen Sie spezifische Verzerreffekte? Hatten Sie dazu eine Weiterbildung?

*P1: Nein, wir haben dazu keine Weiterbildung oder Schulung erhalten, versuchen aber immer jede Prüfung sehr objektiv zu bewerten. Vielleicht sollten wir uns mit der Thematik noch näher auseinandersetzen.*

*P2: Nein, also weder im Studium noch vor den Prüfungen, habe ich etwas zu Verzerreffekten gemacht. Aber bei der ersten richtigen Prüfung hat die Prüferin, ich war Protokollantin, sich sehr viel Zeit genommen, um mit mir über die Prüfung und die Bewertung zu sprechen. Also wirklich dann auch ein „learning by doing“.*

*P3: Nein, auch bei der telc-Qualifizierung wurde das nicht angesprochen. Aber trotzdem haben wir zwischen den Prüfungen Zeit, um darüber zu sprechen, ob wir vielleicht von den schriftlichen Leistungen beeinflusst sind. Und eine meiner Kolleginnen thematisiert auch die möglichen „menschlichen“ Einflüsse auf die Prüfungsbewertung. Aber so ein „Überbau“ zu den Verzerreffekten, also eine Bewusstwerdung, wäre sicher sinnvoll.*

## 3 Fazit und Ausblick

Unabhängig von der Qualität der Prüfungen an sich, die bei *TestDaF* und *telc* standardisiert sind, während die *DSH* nicht standardisiert ist (aber Qualitätsrichtlinien vorliegen), zeigen sich große Unterschiede bei der Prüferqualifizierung. Bei allen drei Institutionen gewinnt man „Prüfungssicherheit“ durch Erfahrung und das Beobachten der Kollegen und Kolleginnen. Eine profunde Qualifizierung erfolgt ausschließlich bei *TestDaF*. Dort zeigt die Befragung der Prüfenden, dass sie ausreichend auf unterschiedliche Facetten vorbereitet werden. Die Zufriedenheit mit der Qualifizierung ist ausgesprochen hoch. Bei der *telc*-Prüferqualifizierung liegen die Vermutungen bezüglich der Integration verschiedener Faktoren durch die Institution selbst und die Bewertung der Prüfenden weiter auseinander. Schwierig ist hier vor allem die sehr heteroge-

ne Vorbildung der „zukünftigen“ Prüfenden. Die Interviewten machten sehr deutlich, dass die Qualifizierung an sich sicherlich erfolgreicher wäre, wenn die Vorbildung der Teilnehmenden ähnlich hoch wäre. Ferner wünschten sich alle Prüferinnen mehr Transparenz und ein ausführlicheres Feedback. Gerade, wenn keine DaF/DaZ-Qualifizierung gefordert ist, wäre zu überlegen, ob man zum einen das Vorwissen besser überprüft und zum anderen die Faktoren der Prüfungsverzerrung in die Qualifizierung integriert. Prüfende bei der *DSH* benötigen zwar inzwischen ein DaF-spezifisches Studium, dies wird aber sicherlich noch nicht an allen Sprachenkollegs umgesetzt sein. Am ISSK Mainz haben jedoch die befragten Prüfenden alle ein spezifisches Studium und/oder sehr viel Erfahrung im Bereich DaF. Ein gezieltes Prüfungstraining fehlt. Die „Qualifizierung“ erfolgt durch Wissensweitergabe der „Prüfungspromis“. Das ISSK Mainz überlegt aufgrund der Interviews, ob eine zusätzliche Weiterbildung nicht sinnvoll wäre.

## Bibliografie

- ALTE (2012): *Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER*, im Auftrag des Europarats/Abteilung für Sprachenpolitik, abrufbar unter: [https://www.telc.net/fileadmin/user\\_upload/handbuch\\_zur\\_entwicklung\\_und\\_durchfuehrung\\_von\\_sprachtests.pdf](https://www.telc.net/fileadmin/user_upload/handbuch_zur_entwicklung_und_durchfuehrung_von_sprachtests.pdf), Mai 2020.
- ALTHAUS, H.-J. (2018): „Warum C1 keine Lösung ist: Der Nachweis von Deutschkenntnissen für den Hochschulzugang, der GER und warum sie nicht zusammenpassen“, in BRANDT, A., BUSCHMANN-GÖBELS, A. & HARSCH, C. (Hrsg.): *Der Gemeinsame Europäische Referenzrahmen für Sprachen und seine Adaption im Hochschulkontext*, Bochum: AKS-Verlag: 78–98.
- ARRAS, U. (2007): „Zur Revision des Subtests ‚Mündlicher Ausdruck‘ der Prüfung ‚Test Deutsch als Fremdsprache‘ (TestDaF)“, in: *Studienkolleg. Zeitschrift zur Pädagogik und Didaktik studienvorbereitender Kurse für ausländische Studierende* 11/2007: 5–29.

- DEMMIG, S. (2012): *Sprachkompetenzen prüfen – aber wie? Zur Konstruktvalidität von standardisierten Sprachtests am Beispiel der DSH: Theoretische Grundlegung und empirische Longitudinalstudie*, Jena: Habilitationsschrift Friedrich-Schiller-Universität Jena.
- ECKES, T. (2003): „Qualitätssicherung beim TestDaF: Konzepte, Methoden, Ergebnisse“, in: *Fremdsprachen und Hochschule* 69: 43–68.
- ECKES, T. (2005): „Evaluation von Beurteilungen: Psychometrische Qualitätssicherung mit dem Multifacetten-Rasch-Modell“, in: *Zeitschrift für Psychologie* 213: 77–96.
- EAQUALS (2019): *Der Equals Rahmen für die Aus- und Weiterbildung von Sprachlehrenden* (Evaluation and Accreditation of Quality in Language Services, dt.: Qualitätsevaluierung und -akkreditierung von Sprachunterricht), abrufbar unter: <https://www.equals.org/our-expertise/teacher-development/the-equals-framework-for-teacher-training-and-development/>, Mai 2020.
- FANDRYCH, C. & MÜLLER, E. (2019): „Deutsch in der Hochschullehre der deutschsprachigen Länder“, in: AMMON, U. & SCHMIDT, G. (Hrsg.): *Förderung der deutschen Sprache weltweit: Vorschläge, Ansätze und Konzepte*: Berlin & Boston: de Gruyter: 291–316.
- GREIN, M. (2019): „Förderung der Sprachkompetenz an Hochschulen: PHILIS – Studieren im Philosophicum“, in: REGIER, K. et al. (Hrsg.): *Innovative Methoden zur Förderung der Sprachkompetenz in der Hochschullehre*, Wiesbaden: Springer: 251–279.
- HOFFMANN, S. (2014): *Mündliche Kompetenz und Bewusstsein beim unterrichtlichen Fremdsprachenlernen*, Tübingen: Narr.
- KECKER, G. (2011): *Validierung von Sprachprüfungen. Die Zuordnung des TestDaF zum Gemeinsamen europäischen Referenzrahmen für Sprachen*, Frankfurt: Peter Lang.
- KECKER, G. (2016): „Was macht eine gute Sprachprüfung aus? Qualitätssicherung beim TestDaF“, in: DRUMBL, H. ET AL. (Hrsg.): *IDT 2013*, Bozen: Bozen-Bolzano University Press: 145–164, abrufbar unter: <https://pro.unibz.it/library/bupress/publications/fulltext/9788860461261.pdf>, Mai 2020.

- KOREIK, U. (Hrsg.) (2005): *DSH und TestDaF – eine Vergleichsstudie*, Baltmannsweiler: Schneider Verlag Hohengehren.
- LENNON, P. (2000): „Die Bewertung mündlicher Sprachkompetenz“, in: BÖRNER, W. & VOGEL, K. (Hrsg.): *Normen im Fremdsprachenunterricht*, Tübingen: Narr: 165–176.
- LUFTFAHRT BUNDESAMT (2019): „Leitlinien für den Erwerb und den Erhalt einer Befähigung zur Abnahme von Sprachprüfungen gem. § 125 LuftPersV“, abrufbar unter: [https://www.lba.de/SharedDocs/Downloads/DE/L/L2/Sprachpruefung/Leitlinien\\_Befaeahigung.pdf?\\_\\_blob=publicationFile&v=4](https://www.lba.de/SharedDocs/Downloads/DE/L/L2/Sprachpruefung/Leitlinien_Befaeahigung.pdf?__blob=publicationFile&v=4), 14.10.2020.
- MARKS, D. (2015) „Prüfen sprachlicher Kompetenzen internationaler Studienanfänger an deutschen Hochschulen: Was leistet der TestDaF?“, in: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 20/1: 21–39.
- MÖHRING, J. & BÄRENFÄNGER, O. (2018): „Hochschulzugangsprüfungen und die Studienrealität: Eine empirische Untersuchung zu Lese- und Wortschatzanforderungen in der Studieneingangsphase“ in: *Info DaF* 2018 45/4: 540–572.
- NORRIS, J. & DRACKERT, A. (2017): „Test review: TestDaF“, in *Language Testing* 35/1: 149–157, abrufbar unter: <https://doi:10.1177/0265532217715848>, 14.10.2020.
- ROCHE, J. (2013): *Fremdspracherwerb Fremdsprachendidaktik*, 3. Auflage, Tübingen: Narr.
- SCHNEIDER, A. (2010): *Zur Standardisierung der DSH-Prüfungen: Bestandsaufnahmen und Perspektiven des Online-Angebotes*, Hamburg: disserta Verlag.
- ŠRÁMKOVÁ, E. (2016): „Beurteilung und Bewertung kommunikativer Leistungen bei mündlichen Prüfungen in der Fachsprache und in der akademischen Sprache“, in DRUMBL, H. et al. (Hrsg.): *IDT 2013*, Bozen: Bozen-Bolzano University Press: 227–232, abrufbar unter: <https://pro.unibz.it/library/bupress/publications/fulltext/9788860461261.pdf>, Mai 2020.
- STEVENER, J. (2018): „Testen und Prüfen in der Fremdsprache: Was macht eine gute Sprachprüfung aus?“, in: *Ramkhamhaeng University Journal Humanities Edition Vol. 37/1*: 21–46, abrufbar unter: <https://so05.tci-thaijo.org/index.php/huru/article/view/133050/99802>, Mai 2020.

- STRAUB, K. (2014): „Kommunikationsstrategien bei mehrsprachigen SchülerInnen im reziprok-immersiven Kindergarten FiBi (Filière Bilingue) in Biel/Bienne“, in: *Babylonia 01/14*: 44–47.
- WOLLERT, M. & ZSCHILL, S. (2017): „Sprachliche Studierfähigkeit: Ein Konstrukt auf dem Prüfstand“, in: *Info DaF 44/2*: 2–17.
- ZEIDLER, B. (2016): „Bewerterkalibrierung bei kleinen Prüfungen“, in DRUMBL, H. ET AL. (Hrsg.): *IDT 2013*, Bozen: Bozen-Bolzano University Press: 203–226, abrufbar unter: <https://pro.unibz.it/library/bupress/publications/fulltext/9788860461261.pdf>, Mai 2020.





## **Sprachliche Heterogenität und deren Einfluss auf das Hörverstehen in Sprachprüfungen – eine Vokalidentifikationsstudie mit Deutschlernenden und Deutschsprechenden**

### **Einleitung**

Hörverstehen stellt für Deutschlernende aufgrund von Faktoren wie der Textart, Aufgabenstellung und nicht zuletzt der Fähigkeiten der Lernenden eine große Herausforderung dar (vgl. Adamczak-Kryzstofowicz & Limbach 2019). Der vorliegende Artikel begutachtet die Vokalidentifikation als basale phonologische Erkennungsleistung von Deutschlernenden im Vergleich zu deutschen Muttersprachlerinnen und Muttersprachlern aus psycholinguistischer Perspektive. Hierfür wurde eine Identifikationsaufgabe mit Probandinnen und Probanden durchgeführt, die entweder monolingual mit Deutsch aufgewachsen sind oder Deutsch mit mindestens einem Sprachniveau von GER B2 erlernt haben. Besonderes Augenmerk wird dabei auf Gruppenunterschiede und Betrachtungen der Varianz der Testergebnisse gelegt. Ziel soll dabei unter anderem sein, den Einfluss zu betrachten, den unterschiedliche Sprachhintergründe der Deutschlernenden auf das Ergebnis der Vokalidentifikation ausüben und daraus Perspektiven für Sprachunterricht und Sprachprüfungen zu entwickeln.

Zu diesem Zweck wird zunächst die Lautverarbeitung bei Sprachlernenden und das Erlernen einer Fremdsprache aus psycho- und neurolinguistischer Perspektive dargelegt (Kapitel 1). Im Anschluss daran werden kurz die Herausforderungen des Hörverstehens in Prüfungen für Deutsch als Fremd- und Zweitsprache aus psycholinguistischer Sicht beleuchtet, um die Motivation der durchgeführten Studie zu verdeutlichen (Kapitel 2). Die vorliegende Studie zur Vokalidentifikation standarddeutscher Langvokale mitsamt den ihr zugrunde-

liegenden Forschungsfragen wird in Kapitel 3 erläutert. Hier werden wichtige Informationen zu den Probandengruppen (3.1) gegeben. Eine Beschreibung des Stimulusmaterials (3.2) sowie der allgemeinen Vorgehensweise folgen in Kapitel 3.3. Daraufhin werden die Daten analysiert und die Ergebnisse der Studie mit Hinblick auf die gewählten Maße dargelegt (3.4). Abschließend werden die Ergebnisse diskutiert und in den Kontext der Sprachprüfungen eingeordnet gefolgt von einem Ausblick im letzten Kapitel für weitere Untersuchungen, der Anregungen zur weiteren Kooperation zwischen Psycho-/Neurolinguistik und Sprachlehre bzw. Sprachprüfungen gibt.

## **1 Psycho- und neurolinguistische Befunde zur Lautverarbeitung bei Sprachlernenden**

Im Rahmen der ersten IDS-Goethe-Studie von 2018, die sich mit Teilnehmenden von Sprachkursen für das Deutsche befasste, wurden Sprachbiografien von 606 Teilnehmerinnen und Teilnehmern aus 42 Integrationskursen erstellt. Die Befragten gaben insgesamt 80 verschiedene Muttersprachen (L1) an, die sie in den ersten fünf Lebensjahren erwarben (vgl. Cindark et al. 2019; Hünlich et al. 2018). Es konnte daher eine beträchtliche Heterogenität sprachlicher Hintergründe beobachtet werden. Daraus folgt, dass in Sprachkursen und Sprachprüfungen auch eine Vielzahl unterschiedlicher Phoneminventare aufeinandertrifft. Entsprechend einer Studie von Best et al. (2001: 789–790) ist davon auszugehen, dass diese muttersprachlichen Phoneminventare jeweils als Basis für den Erwerb deutscher Sprachlaute fungieren. Die Forscherinnen und Forscher konnten zeigen, dass die Fähigkeit, Vokale einer Fremdsprache zu unterscheiden, vom Sprachhintergrund abhängig ist. So wird bei der Perzeption nicht-muttersprachlicher Laute auf das Phoneminventar der Muttersprache zurückgegriffen. Die Versuchspersonen der Studie von Best et al. (2001) konnten in einer Diskriminationsaufgabe Laute besser unterscheiden, wenn die Lautkontraste denen ihrer L1 ähnelten. Nicht verwunderlich sind daher die Ergebnisse einer Untersuchung von Eckes (2013), die implizieren, dass die Erfolgswahrscheinlichkeit bei Hörverstehensaufgaben mit dem individuellen Sprachhintergrund korreliert. Somit liegt nahe, dass der Erwerb von

Deutsch als Fremdsprache (L2<sup>1</sup>) in Abhängigkeit von der Erstsprache der Lernenden unterschiedlich große Herausforderungen birgt. Motiviert durch die evidente Heterogenität in Sprachkursen wird in dieser Studie die Perception fremdsprachlicher Phoneme unter bewusster Einbeziehung von sprachlicher Divergenz und unter Berücksichtigung neuro- und psycholinguistischer Erkenntnisse betrachtet.

Allgemein gilt, dass Kleinkinder ihre Muttersprachen erwerben, indem sie dauerhaft mit ihnen im Kontakt sind. Bis zu einem Alter von etwa sechs Monaten sind Säuglinge prinzipiell in der Lage, Laute aller Sprachen zu perzipieren und zu differenzieren (vgl. Werker & Tees 1984a; Kuhl 2004). Studien zeigen, dass sich das Gehirn im Laufe des ersten Lebensjahres auf die Phoneme der Muttersprache oder Muttersprachen spezialisiert, bei gleichzeitiger Reduktion der Fähigkeit, nicht-muttersprachliche Laute zu identifizieren (vgl. Kuhl et al. 1992; Werker & Tees 1984b).

Nichtsdestoweniger besitzen natürlich auch Erwachsene die Fähigkeit, neue Sprachen zu erlernen. Eine Vielzahl von Studien zur Plastizität des menschlichen Gehirns zeigt, dass auch im Erwachsenenalter neuronale Strukturen angepasst werden können, um den Erwerb einer Fremdsprache zu gewährleisten (vgl. Costa & Sebastián-Gallés 2014; Golestani et al. 2006; Li et al. 2014; Osterhout et al. 2008; Sebastián-Gallés et al. 2012). Mittels bildgebender Verfahren wie der funktionellen Magnetresonanztomographie (fMRT) lässt sich dabei erkennen, dass sich das Gehirn zur Verarbeitung der L2 tendenziell am neuronalen Netzwerk der L1 bedient, also die gleichen Areale und Faserbahnen verwendet werden, wobei jedoch zusätzliche Regionen aktiviert werden müssen (vgl. Marian et al. 2003; Parker Jones et al. 2012; Perani & Abutalebi 2005; Sebastian et al. 2011; Sulpizio et al. 2020). Nicht etwa das Sprachniveau, sondern das Erwerbalters scheint bei der Verarbeitung von Fremdsprachen eine entscheidende Rolle zu spielen. Genauer gesagt sprechen Analysen neuronaler Daten für Aktivierungen in mehr Regionen, wenn der Spracherwerb zu einem späteren Zeitpunkt stattfand (vgl. Gurunandan et al. 2019; Kim et al. 1997; Liu & Cao 2016; Park et al. 2012). In neurolinguistischen Studien wurde

.....

1 Im Weiteren werden alle Fremdsprachen als L2 bezeichnet, ungeachtet der Reihenfolge, in der sie erworben wurden.

zudem der Einfluss der Muttersprache auf das Lernen einer Fremdsprache untersucht. Demnach konnten bei der Bearbeitung derselben Hörverstehensaufgaben, Reimerkennungsaufgaben und lexikalischen Entscheidungsaufgaben abhängig von der L1 der Teilnehmenden Aktivitäten in unterschiedlichen Hirnarealen beobachtet werden (vgl. Joeng et al. 2007; Kim et al. 2016; Yokoyama et al. 2013). Jedoch befassten sich die Forscherinnen und Forscher in erster Linie mit orthografischen und semantischen Fragestellungen, nicht explizit mit dem Einfluss der Distanz zwischen L1 und L2 auf Phonemebene.

Wie Untersuchungen zur L2-Lautverarbeitung bei Erwachsenen mit Hilfe von fMRT zeigen, können bereits bei geringen Fremdsprachkenntnissen Aktivierungsmuster erkannt werden, die der L1-Lautverarbeitung ähneln. So wiesen japanische Muttersprachlerinnen und Muttersprachler mit geringen Englischkenntnissen bei der Verarbeitung von Lauten des Englischen (über die das Japanische nicht verfügt) vermehrt Aktivierungen in Hirnarealen des auditorischen Kortex auf, die auch bei L1-Lautverarbeitung vorhanden waren (vgl. Callan et al. 2004). Ein ähnliches Muster zeichnet sich nach kurzem Lautidentifikationstraining ab, bei dem englischsprachige Versuchspersonen unbekannte Laute des Hindi trainierten. Die Hirnareale waren in dieser Studie vor dem Training nicht aktiviert. Nach dem Training konnten jedoch ebenfalls Aktivierungen gefunden werden, die denen bei L1-Lautverarbeitung ähnelten (vgl. Golestani & Zatorre 2004). Zusätzlich konnte auch in diesen Studien beobachtet werden, dass bei der Perzeption von nicht-muttersprachlichen Lauten mehr Regionen Aktivität aufwiesen als bei muttersprachlichen Lauten. Diese Ergebnisse lassen allerdings nur bedingt Rückschlüsse bezüglich des Einflusses der Ähnlichkeiten der beobachteten Sprachen, des wachsenden Sprachniveaus und dem Zeitpunkt des Spracherwerbs zu.

Archila-Suerte et al. (2012, 2015) argumentieren auf Basis der Ergebnisse ihrer fMRT-Studien, dass nicht etwa das Sprachniveau, sondern insbesondere der Zeitpunkt des Spracherwerbs ausschlaggebend dafür ist, welche Regionen bei der Verarbeitung von L2-Lauten aktiviert werden. So postulieren sie (2015: 46) eine stärkere Aktivierung des *lobus parietalis inferior* zur Identifikation von nicht-muttersprachlichen Lauten bei späten im Vergleich zu frühen Lernenden. Diese Areale werden mit selektiver Aufmerksamkeit assoziiert (vgl. Behrmann et al. 2004). Archila-Suerte et al. (2015: 46) vermuten daher, dass

späte Lernende diese Strategie implementieren, um den Einfluss der Muttersprache auf die Perzeption zu reduzieren und sich somit besser auf die Laute der L2 fokussieren zu können.

Bei Betrachtung der behavioralen Ergebnisse von Archila-Suerte et al. (2012, 2015), Callan et al. (2004) und Golestani & Zatorre (2004) fällt auf, dass Laute einer erlernten Fremdsprache generell schwieriger zu identifizieren sind als die der Muttersprache. Dies stimmt mit zahlreichen weiteren Forschungsbefunden überein, wie z.B. mit denen einer Diskriminationsstudie (vgl. Flege 1995), bei welcher Muttersprachlerinnen und Muttersprachler insgesamt sechs verschiedener Muttersprachen (Arabisch, Deutsch, Koreanisch, Niederländisch, Portugiesisch, Spanisch) Vokale des Englischen unterscheiden mussten. Die Ergebnisse variierten je nach Muttersprache der Teilnehmenden. Hierbei waren zwei englische Vokale dann schwieriger zu unterscheiden, wenn sie durch Assimilation entsprechend als verschiedene Tokens desselben L1-Vokals identifiziert wurden. So fiel bspw. den deutschen Versuchspersonen die Unterscheidung des Kontrasts / $\epsilon$ /-/ $\ae$ / am schwierigsten, da im deutschen Phoneminventar nur der Vokal / $\epsilon$ /, nicht aber / $\ae$ / vorhanden ist. In anderen Worten: Eine Diskrimination von fremdsprachlichen Lauten ist dann erschwert, wenn eine perzeptuelle Nähe zu einem muttersprachlichen Laut besteht oder die fremden Laute sogar innerhalb einer muttersprachlichen Lautkategorie liegen. Dies untermauert besonders den Einfluss der Muttersprache auf die Perzeption von L2-Lauten (vgl. Best & Tyler 2007; Faris 2017; Flege 2002, 2003; Flege et al. 1997; McAllister et al. 2002; Ohala 2008; Strange & Shafer 2008). Auch die Annahme, dass Laute der erlernten Sprache eher in Richtung der muttersprachlichen Phoneme assimiliert werden, wenn sie diesen ähneln (vgl. Best et al. 2001), kann bestätigt werden. Ein Experiment, dessen Teilnehmerinnen und Teilnehmer dialektale Unterschiede aufweisen, berichtet ähnliche Resultate. Während einer Kategorisierungsaufgabe hörten Sprecherinnen und Sprecher des holländischen und des westflämischen Dialekts der niederländischen Sprache die englischen Vokale / $\epsilon$ / und / $\ae$ /, welche sie Phonemen des britischen Englisch zuordnen mussten. Die Dialektgruppen unterschieden sich in der Perzeption der fremdsprachlichen Vokale, wobei Sprecherinnen und Sprecher des holländischen Dialekts / $\epsilon$ / häufiger und / $\ae$ / seltener korrekt identifizierten als die westflämische Gruppe. Dies zeigt,

dass sich dialektale Differenzen ebenfalls in den Identifikationsleistungen der L2-Laute widerspiegeln (vgl. Escudero et al. 2012).

Die von Archila-Suerte et al. (2012, 2015) formulierten Thesen zum Einfluss des Zeitpunkts des Spracherwerbs decken sich ebenfalls mit einer Studie von Flege & MacKay (2004). Dies geht aus dem Abschneiden zweier Gruppen italienischer Muttersprachlerinnen und Muttersprachler hervor, die im Schnitt seit drei Monaten oder seit 36 Jahren kanadisches Englisch sprachen (vgl. Flege & MacKay 2004). Bei der Diskrimination von L2-Vokalkontrasten schnitten die Spracherfahrenen besser ab als die Sprachunerfahrenen. Darüber hinaus erzielten Spracherfahrene, die angaben, ihre Muttersprache häufig zu verwenden, signifikant schlechtere Ergebnisse als diejenigen, die angaben, selten Italienisch zu sprechen. Folglich hat nicht nur die Länge des Aufenthalts und die Intensität des Sprachenlernens, sondern auch die Häufigkeit des Gebrauchs der L1 einen Einfluss auf die Perzeption von nicht-muttersprachlichen Vokalen.

Schon Flege et al. (1997) fokussierten die Frage, inwiefern Spracherfahrung und Identifikationsleistung korrelieren. Im Rahmen einer Identifikationsaufgabe mit Auswahlmöglichkeiten waren Muttersprachlerinnen und Muttersprachler des Deutschen, Spanischen und Koreanischen mit unterschiedlich guten Englischkenntnissen angewiesen, die englischen Vokale der Kontinua *beat-bit (/i/-/ɪ/)* und *bat-bet (/ɛ/-/æ/)* zu identifizieren. Die Teilnehmenden sprachen außer Englisch keine weitere Fremdsprache. Neben Einflüssen der unterschiedlichen Muttersprachen fanden Flege et al. (1997) heraus, dass sich mit höherem L2-Sprachniveau auch die Identifikation der L2-Vokale verbessert. Zu ähnlichen Schlüssen kamen Bundgaard-Nielsen et al. (2011), wobei sie hervorheben, dass ein größeres Vokabular in der Fremdsprache mit konstanterer Assimilation und besserer Diskrimination der L2-Vokale korreliert. Jedoch scheint der Effekt des wachsenden Vokabulars auf die Diskriminationsleistung ab einer Größe von etwa 6000 Wörtern nicht mehr zum Tragen zu kommen.

Im Kontext von Sprachprüfungen und Sprachunterricht ist außerdem der Zusammenhang zwischen Produktion und Perzeption von L2-Lauten relevant. Studien belegen, dass Vokale differenzierter produziert werden, wenn sie auch genauer identifiziert und unterschieden werden (vgl. Flege et al. 1997, Perkell et al. 2004), d.h. das korrekte Hörverstehen hat Einfluss auf die Aussprache, in diesem Falle auf die Aussprache der Vokale.

Insgesamt heben die Beobachtungen hervor, wie essenziell der Faktor Zeit für den erfolgreichen Erwerb des Phoneminventars einer Zweitsprache ist. Kompetenzzuwachs und neuronale Restrukturierung der phonologisch relevanten Hirnareale scheinen hierbei Hand in Hand zu gehen. Zusätzlich zeigt sich, dass für erfolgreiches Hörverstehen basale perzeptive Prozesse notwendig sind. So bildet beispielsweise die phonologische Lauterkennung die Grundvoraussetzung für erfolgreiche Sprachperzeption. Darüber hinaus wurde auch die enge Verbindung von Sprachperzeption und -produktion veranschaulicht. Besonders deutlich wird jedoch in der Zusammenschau, dass aus psycho- und neurolinguistischer Perspektive der individuelle Sprachhintergrund einen der einflussreichsten Faktoren beim Fremdsprachenlernen darstellt, der in Sprachprüfungen bislang eher vernachlässigt wird. Welche Rolle das Hörverstehen in Deutsch als Fremdsprache-Prüfungen genauer einnimmt, wird im folgenden Kapitel skizziert.

## **2 Die Fertigkeit Hören in Prüfungen für Deutsch als Fremd- und Zweitsprache: Herausforderungen aus psycholinguistischer Sicht**

Das Studium an einer deutschen Hochschule setzt nachweisbare Deutschsprachkenntnisse auf dem Niveau C1 voraus. Ein solcher Nachweis kann erworben werden durch das erfolgreiche Absolvieren einer Sprachprüfung wie zum Beispiel *TestDaF*, *DSH* und *telc Deutsch C1 Hochschule*. Bestandteil dieser Prüfungen sind neben Leseverstehensaufgaben, der Produktion von schriftlichen Texten und dem mündlichen Prüfungsteil auch Aufgaben, die das Hörverstehen anhand von auditiv präsentierten Texten abfragen. Nach dem Gemeinsamen Europäischen Referenzrahmen für Sprachen bzw. dem neueren Begleitband von 2018 werden die globalen Fähigkeiten im Bereich Hörverstehen auf C1-Niveau wie folgt definiert:

Can understand enough to follow extended speech on abstract and complex topics beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar.



Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts.

Can follow extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. (Council of Europe 2018: 55)

Die Prüfungen setzen demnach insbesondere voraus, dass die Teilnehmenden über ein umfangreiches Vokabular verfügen und vermögen, Inhalte auf Basis ihres Kontexts zu verknüpfen. Diesen Fähigkeiten liegt zugrunde, dass die Geprüften sich der Phonologie des Deutschen bewusst sind und Wörter in gesprochener Sprache anhand ihrer Lautstruktur identifizieren können (vgl. Böttger 2016: 129). Auf einem hohen Sprachniveau ist von einem großen Vokabular und damit korrelierend von guten perzeptiven Kompetenzen auszugehen (vgl. Bundgaard-Nielsen et al. 2011). Nichtsdestoweniger geben die Ergebnisse der in Kapitel 1 vorgestellten neuro- und psycholinguistischen Studien Anlass zur Vermutung, dass der Sprachhintergrund der Teilnehmenden einen Einfluss auf deren Abschneiden im Bereich Hörverstehen hat. So lässt sich vermuten, dass weniger Überschneidungen des Phoneminventars von Muttersprache und erlernter Fremdsprache mit einem erhöhten Lernaufwand bezüglich der Vorbereitung auf Hörverstehensaufgaben einhergehen. Dies wirft die Frage auf, ob die Teilnehmenden bei Sprachprüfungen abhängig von ihrem individuellen sprachlichen Hintergrund die gleichen Chancen auf Erfolg haben bzw. Sprachprüfungen *per se* fair sein können.

Am Beispiel der Qualitätssicherung des *TestDaF*<sup>2</sup> zeigt sich bspw. ein ständiges Bestreben, Sprachtests zu optimieren und deren Validität zu gewährleisten. Testentwicklung, Durchführung und Bewertung werden dabei für jede neu entwickelte Sprachprüfung empirisch erprobt (vgl. Eckes 2003, 2005, 2008; Kecker 2016). Dazu werden unter anderem Methoden entworfen, die ermöglichen sollen, dass alle Teilnehmenden trotz Einschränkungen, wie zum Beispiel Blindheit, möglichst gleiche Chancen auf erfolgreiches Bestehen des *TestDaF* haben (vgl. Arras et al. 2013). Des Weiteren wird der Schwierigkeitsgrad jedes Tests so angepasst, dass die Aussagekraft der

.....  
2 Genauerer dazu im Beitrag von Grein in diesem Band.

Sprachniveaustufen konstant bleibt. Auch Variationen in der Bewertung, die aus der Strenge oder Milde der Prüferinnen und Prüfer resultieren, unterliegen stetiger Erprobung. Dadurch soll sichergestellt werden, dass tatsächlich die reine sprachliche Leistung der Teilnehmenden bewertet wird (vgl. Eckes 2005).

Doch aus diesem Grund gestaltet sich die Berücksichtigung des Sprachhintergrundes bei der Bewertung komplex. Bedingt durch die sprachliche Heterogenität in Sprachkursen herrscht ein Ungleichgewicht bezüglich der Intensität der Vorbereitung für Sprachprüfungen. Da in Sprachprüfungen jedoch das Sprachniveau beurteilt wird, nicht aber der Lernaufwand, könnten Abweichungen der Evaluation, die aufgrund individueller Spracherfahrungen der Teilnehmenden erfolgen, die Aussagekraft des Sprachtests schmälern (vgl. Plassmann & Zeidler 2014: 239).

Dieser Argumentation stehen nicht nur die Erkenntnisse aus psycho- und neurolinguistischen Studien gegenüber. Eckes (2013: 378) konnte bei einer Analyse der Prüfungsergebnisse des *TestDaF* zeigen, dass das Abschneiden der Prüflinge bei Hörverstehensaufgaben auf den Sprachhintergrund zurückzuführen ist. In seinen Resultaten fasst der Autor die Teilnehmenden zu vier Gruppen zusammen: chinesische, koreanische, slawische und westeuropäische Muttersprachlerinnen und Muttersprachler. Zwar ist diese Unterteilung typologisch unpräzise, dennoch konnten Unterschiede bezüglich des Hörverstehens des Deutschen gefunden werden. Hierbei waren die Differenzen am ausgeprägtesten zwischen der chinesischen und slawischen Gruppe sowie zwischen der chinesischen und westeuropäischen Gruppe. Diese Tendenzen bestätigten sich in der Untersuchung eines weiteren *TestDaF*, der zu einem anderen Zeitpunkt durchgeführt wurde und dementsprechend andere Fragen beinhaltete. Dass in beiden Prüfungen Unterschiede bestehen, diese aber in der Stärke ihrer Ausprägung divergieren, macht deutlich, dass der Sprachhintergrund die Leistung beeinflusst, was jedoch in der Prüfungskonzeption des *TestDaF* nicht kontrolliert wird. In Anbetracht der großen sprachlichen Heterogenität bei Sprachprüfungen könnten die Ergebnisse folglich dafür sprechen, den Faktor Sprachhintergrund bereits bei der Prüfungskonzeption miteinzubeziehen.

Ebenso wie die Entwicklung einer Sprachprüfung kann die Lehre einen entscheidenden Einfluss auf die Chancengleichheit haben. Das wachsende Feld der Neurodidaktik untersucht unter anderem Strategien, wie der Sprachunterricht von psycho- und neurolinguistischem Wissen profitieren kann (siehe bspw. Grein et al. 2020). Aktuell sehen die Lehrbücher für das Erreichen des C1-Niveaus hinsichtlich der Vorbereitung auf Hörverstehensaufgaben in erster Linie Textverstehensübungen vor. Dies zeichnet sich zwar durch Prüfungsnähe aus, vernachlässigt aber das Training der phonologischen Grundkenntnisse (vgl. Marks 2015: 30). Wie bereits erwähnt, weist vieles darauf hin, dass die Hörverstehensleistungen von zunehmender phonologischer Kompetenz in der Fremdsprache profitieren (s. Kapitel 1). Ansätze der Neurodidaktik empfehlen daher, zusätzlich Übungen in die Lehre zu inkorporieren, die das Unterscheiden und Identifizieren von nicht-muttersprachlichen Sprachlauten trainieren (vgl. Böttger 2016: 127ff.). Ziel dieser Übungen ist es, die basalen phonologischen Fähigkeiten zu festigen, um schließlich auch die sprachliche Heterogenität eines Sprachkurses auszugleichen. Somit könnte eine Symbiose von Sprachlernforschung und Neurodidaktik das Problem der Chancengleichheit aufgrund unterschiedlicher Sprachhintergründe bereits in der Lehre minimieren.

Obwohl die Vermutung naheliegt, dass die diskutierten Ansätze das Testgütekriterium Fairness von Sprachprüfungen steigern, bedarf es noch einer Vielzahl empirischer Belege, damit die Optimierung von Prüfung und Lehre adäquat gestaltet werden kann. Die Durchführung einer Lautidentifikationsaufgabe (vgl. Dorman et al. 1985; Fry et al. 1962) mit einer sprachlich heterogenen Gruppe erfahrener Deutschlernender eignet sich, um dies realitätsnah zu überprüfen. Da die Ergebnisse psycho- und neurolinguistischer Studien darauf hindeuten, dass selbst Deutschlernende mit hohem Sprachniveau größere Schwierigkeiten bei der Verarbeitung von deutschen Lauten aufweisen als Muttersprachlerinnen und Muttersprachler des Deutschen, sollten letztere als Vergleichsgruppe einer Studie herangezogen werden, die im Folgenden vorgestellt wird.

### 3 Identifikationsstudie

In der vorliegenden Studie wurde ein Identifikationstest mit sechs deutschen Langvokalen durchgeführt. Im Vergleich zu einer Diskriminationsaufgabe (vgl. Gerrits & Schouten 2004; Norris et al. 2003; Polka & Bohn 2003), welche den Fokus auf das Unterscheiden von kontrastierenden Lauten legt, kann davon ausgegangen werden, dass das Testen der reinen Identifikationsleistung basale phonologische Fertigkeiten widerspiegelt (vgl. Fry et al. 1962; Schouten & van Hessen 1992). Die Aufgabe der Versuchspersonen bei einem Identifikationstest ist die Benennung der gehörten Vokale. Dadurch kann beobachtet werden, mit welchen phonologischen Repräsentationen die akustischen Signale der deutschen Vokale verknüpft sind.

Die Anlage der hier vorliegenden Identifikationsstudie ist im psycholinguistischen Sinne insofern ungewöhnlich, als dass der Sprachhintergrund der Sprachlernenden nicht kontrolliert wurde. Mit Kontrolle ist in diesem Zusammenhang gemeint, dass nicht gezielt Sprachlernende einer Sprache oder Sprachfamilie rekrutiert wurden, sondern bewusst eine heterogene Probandengruppe gewählt wurde. Damit soll die Situation der Lehre und Prüfung für Deutsch als Fremd- oder Zweitsprache in der Untersuchung repräsentiert werden, da auch die Lern- und Prüfgruppen heterogene Muttersprachen besitzen (s. Ergebnisse der Goethe Studie, Kapitel 1).

Auch deshalb ist bei dieser Studie der Vergleich von Deutschlernenden und Muttersprachlerinnen und Muttersprachlern interessant. Er erlaubt es, zu untersuchen, inwieweit sich die beiden Probandengruppen voneinander unterscheiden. Die beiden Maße, Akkuratheit und Fehlerart, sollen darüber hinaus auch Aufschluss darüber geben, ob es innerhalb der Untersuchungsgruppen generell schwieriger zu identifizierende Vokale gibt. Im Gruppenvergleich können zusätzlich noch Erkenntnisse darüber gewonnen werden, ob bestehende Schwierigkeiten eher universeller Natur sind oder ob sie mit der Spracherfahrung in Zusammenhang stehen. Dementsprechend lassen sich die Untersuchungsziele zweigeteilt betrachten:

1. Es soll beobachtet werden, ob die Sprachlernenden generelle Schwierigkeiten mit der Identifikation von bestimmten deutschen

Langvokalen haben. Zur besseren Einschätzung der Effekte wird zusätzlich eine Kontrollgruppe aus monolingualen Deutschsprechenden erhoben. Durch den Gruppenvergleich können auch universell schwieriger oder einfacher zu identifizierende Vokale aufgedeckt werden.

2. Durch die im psycholinguistischen Sinne ungewöhnliche Heterogenität der Sprachlernenden (vor allem in Bezug auf den Sprachhintergrund) soll untersucht werden, ob sich diese auch in den Ergebnissen widerspiegelt (z.B. hohe Varianz der Akkuratheit, Varianz der Fehlertypen).

### 3.1 Teilnehmende

Getestet wurden insgesamt 45 Studentinnen und Studenten sowie Mitarbeiterinnen und Mitarbeiter (34 weiblich, 11 männlich) der Johannes Gutenberg-Universität Mainz. Davon bildeten 37 (27 weiblich, 10 männlich) die Gruppe der monolingual Deutschsprachigen und sieben (6 weiblich, 1 männlich) die Gruppe der Deutschlernenden. Im Durchschnitt waren die Teilnehmenden über beide Gruppen hinweg 25,52 Jahre (SD 4,92), die Muttersprachlerinnen und Muttersprachler des Deutschen 25,16 Jahre (SD 4,97) und die Deutschlernenden 27,43 Jahre (SD 4,43) alt.

Für die Gruppe der Muttersprachlerinnen und Muttersprachler galt die Voraussetzung, dass die Teilnehmerinnen und Teilnehmer monolingual mit Deutsch sozialisiert wurden. Für die Gruppe der Deutschlernenden wurde ebenfalls ein monolinguales Aufwachsen mit einer Muttersprache und darüber hinaus ein Deutsch-Sprachniveau von mindestens GER B2 vorausgesetzt. Von den sieben Deutschlernenden waren sechs nach Selbstauskunft auf Niveau C1 und eine Deutschlernende auf Niveau B2. Die Muttersprachen der Deutschlernenden waren Arabisch, Serbokroatisch, Usbekisch, Vietnamesisch, Koreanisch, Rumänisch und Chinesisch. Neben Deutsch waren bei ihnen am häufigsten die Fremdsprachen Englisch (7) und Russisch (3) vertreten. Die monolingual Deutschsprachigen verfügten über Kenntnisse in Englisch (37), Französisch (22), Spanisch (12), Schwedisch (6), Arabisch (4), Italienisch (3), Finnisch (3) und 12 weiteren Sprachen.

### 3.2 Material

Für die Erhebung wurden je fünf Tokens der folgenden sechs deutschen Langvokale verwendet (Beispielwörter jeweils in Klammer): /o:/ (Boot), /u:/ (Hut), /ɛ:/ (Bär), /e:/ (See), /ø:/ (Löwe) und /y:/ (Süd).

Bei Kontrolle der Auftrettsfrequenzen dieser deutschen Langvokale mit Hilfe der *CELEX* Datenbank (vgl. Baayen et al. 1995) konnten Aichert et al. (2005) folgende Rangfolge der hier gebrauchten deutschen Langvokale aufstellen: Der Vokal /e:/ tritt im gewählten Korpus mit insgesamt 84.902 Erscheinungen pro Million Wörter am häufigsten auf, gefolgt von /o:/ (52.854), /u:/ (38.132), /y:/ (21.076), /ɛ:/ (11.295) und /ø:/ (7.542). Die Kontrolle der Auftrettsfrequenz erfolgt, um den Effekt der Frequenz auf die Identifikationsergebnisse einschätzen zu können.

Da von jedem der sechs Langvokale fünf Tokens als Experimentalstimuli verwendeten wurden, beläuft sich die Gesamtzahl auf 30 Experimentalstimuli. Die Stimuli für die hier beschriebene Identifikationsaufgabe dienten bereits in zwei vorangehenden psycho- und neurolinguistischen Experimenten als Stimulusmaterial. Deshalb wurden die Langvokale von zwei phonetisch geschulten Sprecherinnen des Standarddeutschen eingesprochen, wobei eine Sprecherin die Laute /ɛ:/, /ø:/ und /y:/ und die andere Sprecherin die Laute /o:/, /u:/ und /e:/ einsprach. Die Aufnahmen erfolgten im Phonetiklabor der Philipps-Universität Marburg und im Neurolinguistics Lab der Johannes Gutenberg-Universität Mainz, jeweils in einer schallisolierten Kabine. Die Langvokale wurden isoliert und mit einer neutralen Intonation von den Sprecherinnen produziert, um Koartikulations- und Intonationseffekte auszuschließen. Approximativ produzierten die Sprecherinnen insgesamt circa 40 Tokens pro Vokalkategorie.

Im Anschluss daran wurden alle Vokalaufnahmen auf eine Länge von 400 ms mit einem Fade-Out von 50 ms mittels der Software *Praat* (vgl. Boersma & Weenink 2019) geschnitten. Ihre Lautstärke wurde auf 70 dB im Mittel normalisiert. Die ersten und zweiten Formanten (F1, F2) aller Aufnahmen wurden kontrolliert. Es wurden die Aufnahmen für das Experiment ausgewählt, die den stabilsten Höreindruck hinterließen und deren Formanten bei

geringer Varianz am geringsten von den Werten für standarddeutsche Vokale von Sendlmeier & Seebode (2006) abwichen (Abb. 1).

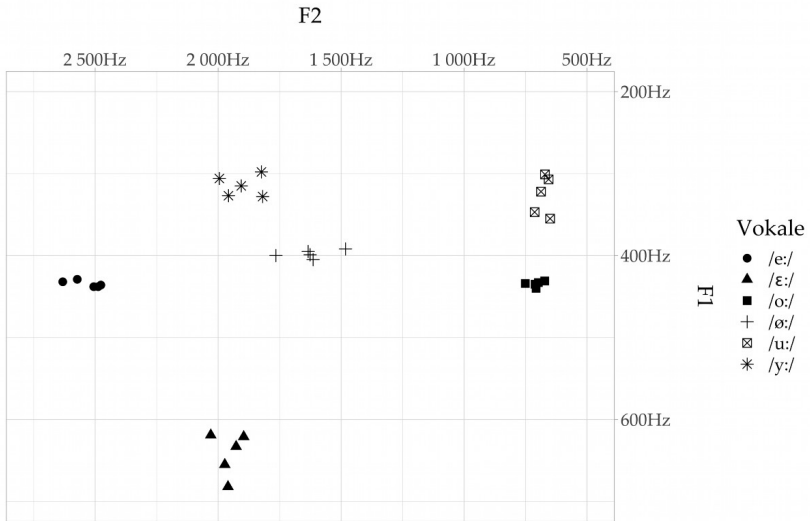


Abb. 1: Formantkarte der Stimuli. Darstellung der Lage der Vokale im Vokalraum anhand des ersten Formanten (F1) und des zweiten Formanten (F2). Vokalkategorien sind ohne Überlappungen zu erkennen

### 3.3 Vorgehen

Die Stimuli wurden den Probandinnen und Probanden jeweils zwei Mal über Lautsprecher in pseudo-randomisierter Reihenfolge präsentiert, sodass sie insgesamt 60 Vokale hörten. Kein Vokal trat mehr als zwei Mal hintereinander auf, um Prädiktions- und Gewöhnungseffekte zu vermeiden. Die Testungen waren in vier Blöcke unterteilt, welche aus je 15 Aufnahmen bestanden. Um zusätzliche Reihenfolgeeffekte auszuschließen, wurden drei Listen erstellt, die in der Präsentationsreihenfolge der Blöcke variierten. Letztere wurden auf die Gruppen aufgeteilt. Nach der Hälfte der Stimuli wurde eine kurze Pause eingelegt.

Durchgeführt wurde das Experiment in Gruppen von sechs bis zehn Versuchspersonen, die im Neurolinguistics Lab der Johannes Gutenberg-Uni-

versität Mainz gleichzeitig an der Paper-Pencil-Befragung teilnahmen. Die Versuchsleitung gewährleistete, dass die Teilnehmenden die Aufgabenstellung selbstständig und für sich allein lösten. Zusätzlich wurde sichergestellt, dass abgegebene Antworten nachträglich nicht mehr korrigiert wurden. Die Vokale wurden in einem langsamen Tempo nacheinander von der Versuchsleitung abgespielt, wobei darauf geachtet wurde, dass die Teilnehmenden genügend Zeit zum Notieren der jeweiligen Antworten hatten.

Die Aufgabe der Probandinnen und Probanden war es, die gehörten Laute schriftlich festzuhalten. Damit die Teilnehmenden sich während des Experiments vollends auf die Identifikation der Laute konzentrieren konnten, wurden sie gebeten, auf eine phonemische Schreibweise zu verzichten und stattdessen auf die deutsche Orthografie zurückzugreifen. Vorgabe war hierbei eine generische Graphem-Phonem-Entsprechung, die ohne genauere Erläuterung von allen Versuchspersonen verstanden wurde: /o:/ = <o>, /u:/ = <u>, /ɛ:/ = <ä>, /e:/ = <e>, /ø:/ = <ö> und /y:/ = <ü>. Abgesehen von dieser Instruktion hatten die Teilnehmer und Teilnehmerinnen keine Information darüber, welche Laute sie hören würden.

### 3.4 Ergebnisse

Wie bereits zuvor erwähnt, wurden in der Datenauswertung zwei Maße betrachtet. Zum einen erfolgte eine Analyse der Akkuratheit (Prozent richtige Antworten) und zum anderen eine qualitative Untersuchung der Fehlerarten (welcher Laut wurde statt des richtigen identifiziert).

#### 3.4.1 Akkuratheit

Zunächst wurden die Antworten der Teilnehmerinnen und Teilnehmer ausgewertet. Für jede Versuchsperson wurde vermerkt, wie viele Aufnahmen sie pro Vokal richtig erkannt hatte. Auf Basis dessen konnte errechnet werden, zu wie viel Prozent die Laute im Durchschnitt korrekt zugeordnet wurden. Im Folgenden wird das arithmetische Mittel in Prozent ( $\emptyset$ ) als Maß der zentralen Tendenz untersucht. Dabei kann das Niveau des Mittelwertes Aussagen über die Güte der gemittelten Antworten zulassen. Hintergrund ist die Annahme, dass bei einem Prozentmittel von circa 50% davon ausgegangen werden muss, dass der Befund



nicht mehr systematisch, sondern zufällig ist. Zusätzlich zum Mittel wird auch die Standardabweichung in Prozent (SD) als Maß der Variabilität betrachtet.

Wie in Tabelle 1 zu erkennen ist, konnten die Langvokale /o:/, /ɛ:/, /e:/ und /ø:/ über beide Gruppen hinweg zu mindestens 92,95% erkannt werden. Der mittlere Vokal /o:/ scheint mit einer Identifikationsrate von 99,55% (SD 2,11%) der Laut zu sein, der am fehlerfreiesten war. Die hohen Langvokale /u:/ und /y:/ hingegen weisen deutlich niedrigere Prozentwerte auf. Im Durchschnitt wurde /u:/ zu 77,05% (SD 28,66) und /y:/ zu nur 67,73% (SD 34,23) identifiziert. Die hohen Standardabweichungen der zuletzt genannten Vokale im Vergleich zu /o:/ deuten darauf hin, dass dies auf Varianz zwischen den Versuchspersonen, aber auch zwischen den Gruppen zurückzuführen sein könnte. Die Standardabweichungen der Vokale /ɛ:/, /e:/ und /ø:/ weisen eine Streuung von mindestens 11,53% auf und haben somit ebenfalls eine höhere Varianz als /o:/.

GESAMT	/o:/	/u:/	/ɛ:/	/e:/	/ø:/	/y:/
Ø (korrekt in %)	99,55	77,05	96,14	97,05	92,95	67,73
SD (korrekt in %)	2,11	28,66	12,05	11,53	20,53	34,23

Tabelle 1: Durchschnittliche Identifikationsleistungen (Ø) und Standardabweichungen (SD) über beide Gruppen hinweg

Um die beiden Forschungsfragen (universell einfachere Vokale, Heterogenität der Ergebnisse als Folge der heterogenen Gruppen) beantworten zu können, werden in einem nächsten Schritt die Werte der beiden Gruppen (Muttersprachlerinnen/Muttersprachler, Deutschlernende) getrennt voneinander betrachtet. Abbildung 2 veranschaulicht diesbezüglich nicht nur, dass die Vokale von den Sprachgruppen unterschiedlich oft passend notiert wurden, sondern auch, dass die Muttersprachlerinnen und Muttersprachler des Deutschen konstant besser abschnitten als die Deutschlernenden. Letztere ordneten im Mittel 45,14 der 60 Vokale (75,24%) richtig zu, wobei die Gesamtzahl der richtigen Antworten pro Versuchsperson von 37 bis 58 reichte. Im Vergleich dazu identifizierten die Monolingual Deutschsprachigen im Durchschnitt 54,54 von 60 Langvokalen, was 90,90% der Stimuli entspricht. Die Anzahl der korrekt erkannten Vokale reichte von 40 bis 60.

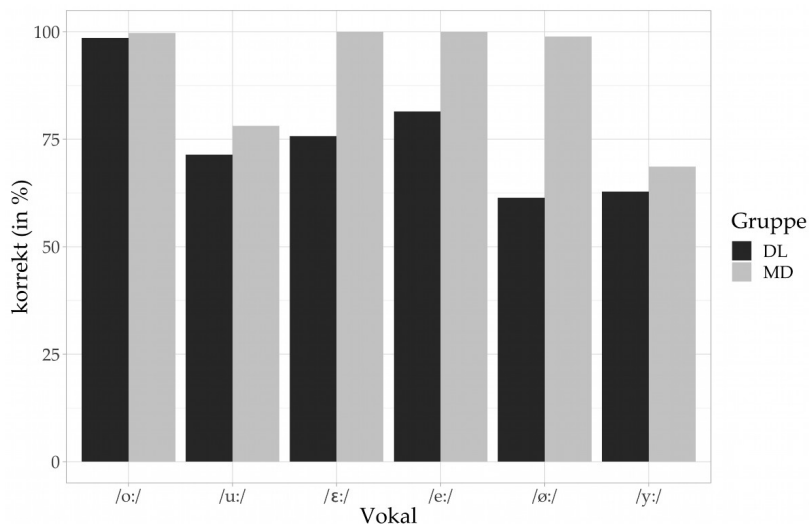


Abb. 2: Gruppenvergleich als Balkendiagramm. Prozentwerte der korrekten Antworten, getrennt nach Vokal und Gruppe (MD = Monolingual Deutschsprachige, DL = Deutschlernende)

Genauer betrachtet zeigen die Ergebnisse der monolingual Deutschsprachigen einen Deckeneffekt bei den Lauten /o:/, /ɛ:/, /e:/ und /ø:/, welche allesamt zu mindestens 98,92% korrekt erkannt wurden (Tabelle 2). Die Standardabweichung liegt hierbei stets unter 3,15%. Zudem wurden die Langvokale /ɛ:/ und /e:/ ausnahmslos fehlerfrei identifiziert. Wie sich ebenfalls in Tabelle 2 abzeichnet, hatten die Muttersprachlerinnen und Muttersprachler Probleme, die Laute /u:/ und /y:/ korrekt zu identifizieren. Der Mittelwert der Identifikation liegt für /u:/ bei 78,11%, für /y:/ bei 68,65%. Die Standardabweichungen von 29,80% für /u:/ und 32,50% für /y:/ spiegeln wider, was ebenfalls in Abbildung 3 zu erkennen ist: Innerhalb der Gruppe der monolingual Deutschsprachigen weichen die Identifikationsleistungen bezüglich der beiden Vokale zwischen den Teilnehmenden stark voneinander ab. Hier kann also konstatiert werden, dass scheinbar auch monolinguale deutsche Muttersprachler eine perzeptuelle Schwierigkeit mit den beiden Randvokalen /u:/ und /y:/ aufweisen, welche jedoch zwischen den Teilnehmenden zu schwanken scheint. Zwar wurde der hintere Langvokal /u:/ von 17 der 37 deutschen Teilnehmerinnen und Teilneh-

mern fehlerfrei erkannt, jedoch schafften dies sieben Versuchspersonen nur zu 50% oder weniger. In zwei Fragebögen wurde der Vokal durchweg nicht notiert, somit wurde hier /u:/ gar nicht erkannt. Den vorderen Langvokal /y:/ identifizierten nur 11 Probandinnen und Probanden zu 100%. Mehr als ein Drittel der monolingual Deutschsprachigen (13 von 37) hörten den korrekten Laut in höchstens 50% der Fälle, wiederum zwei von ihnen erkannten ihn nie. Es konnte keine Korrelation zwischen den beherrschten Fremdsprachen und der Identifikationsleistung festgestellt werden.

MD	/o:/	/u:/	/ɛ:/	/e:/	/ø:/	/y:/	Total
Ø (korrekt in %)	99,72	78,11	100	100	98,92	68,65	90,90
SD (korrekt in %)	1,64	29,80	0	0	3,15	32,50	8,92

Tabelle 2: Durchschnittliche Identifikationsleistungen (Ø) und Standardabweichungen (SD) der monolingual Deutschsprachigen (MD)

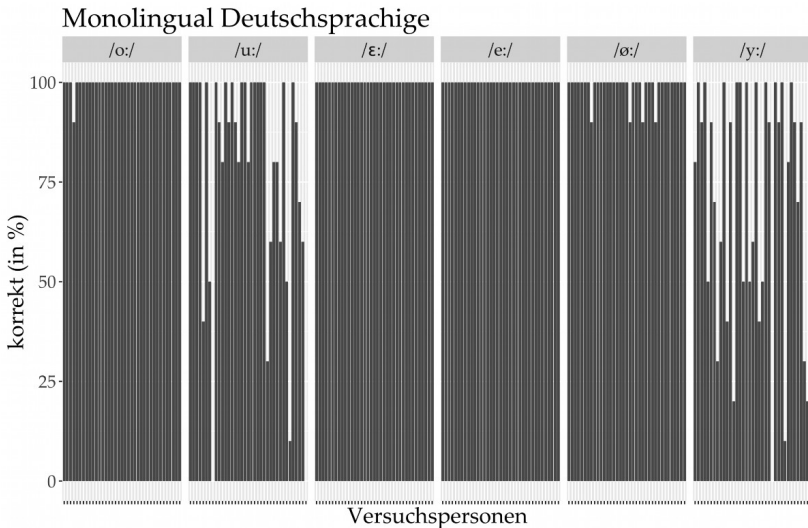


Abb. 3: Identifikationsleistungen (in %) der monolingual Deutschsprachigen, aufgeteilt nach Vokal und Versuchsperson

Ähnlich der Kontrollgruppe konnten die Deutschlernenden /o:/ zu 98,57% (SD 3,78%) weitgehend problemlos benennen (Tabelle 3). Während die deutschen Muttersprachler und Muttersprachlerinnen beim mittleren Langvokal /e:/ und beim Vokal /ɛ:/ einen Deckeneffekt aufwiesen, zeigten die Deutschlernenden hier etwas schlechtere Identifikationsleistungen. Der Langvokal /e:/ wurde in 81,43% (SD 24,78), /ɛ:/ in nur 75,14% (SD 21,49) der Fälle erkannt. Ähnlich schwierig gestaltete sich die Identifizierung des Vokals /u:/. Die Deutschlernenden konnten hier nur in 71,43% (SD 22,68) der Fälle den Vokal richtig identifizieren. Auffällig ist an dieser Stelle, dass in der Gruppe der Deutschlernenden /o:/ ebenfalls der am einfachsten auszumachende Vokal zu sein scheint. Die Vokale /e:/, /ɛ:/ und /u:/ wurden im Vergleich dazu schlechter erkannt. Dementsprechend zeigten die Deutschlernenden hier auch eine größere Varianz der Leistung.

DL	/o:/	/u:/	/ɛ:/	/e:/	/ø:/	/y:/	Total
Ø (korrekt in %)	98,57	71,43	75,14	81,43	61,43	62,86	75,24
SD (korrekt in %)	3,78	22,68	21,49	24,78	39,76	44,99	11,52

Tabelle 3: Durchschnittliche Identifikationsleistungen (Ø) und Standardabweichungen (SD) der Deutschlernenden (DL)

Die Umlaute /ø:/ und /y:/ weisen mit 61,43% respektive 62,86% die schlechtesten Identifikationsleistungen auf. Unter der Annahme eines Zufallsniveaus der Antworten bei circa 50% sind diese Ergebnisse ebenfalls auffällig. Darüber hinaus lässt sich auch hier eine hohe Varianz der Erkennungsleistung zwischen den Teilnehmenden erkennen (Abb. 4). Die Standardabweichungen liegen bei 39,76% für /ø:/ und 44,99% für /y:/. Drei Versuchspersonen identifizierten den Langvokal /ø:/ nur in höchstens der Hälfte der Fälle. Von einer dieser drei Versuchspersonen wurde er zu 0% erkannt. Als Muttersprachen gaben sie Arabisch (Damaskus), Serbokroatisch und Chinesisch (Chengdu) an. Der Langvokal /y:/ wurde von zwei Deutschlernenden konstant nicht notiert. Es handelt sich hierbei um Muttersprachlerinnen und Muttersprachler des Rumänischen und des Koreanischen. Die restlichen fünf Teilnehmerinnen und Teilnehmer identifizierten den hohen Langvokal zu mindestens 50%.

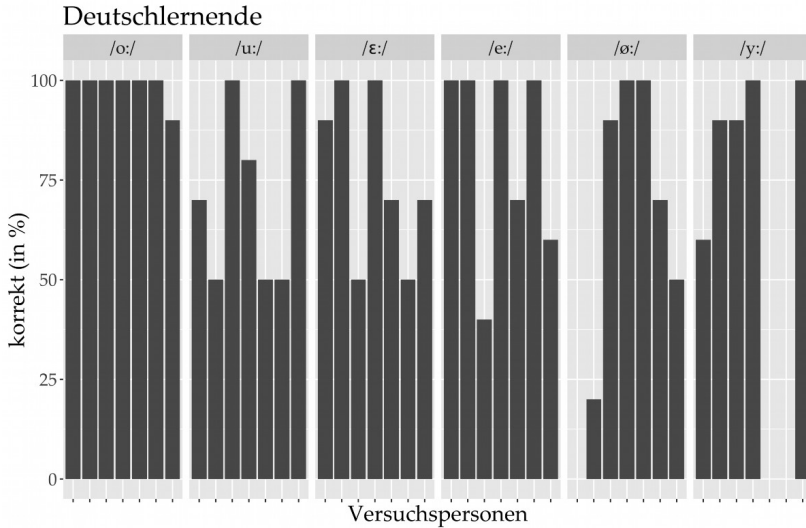


Abb. 4: Identifikationsleistungen (in %) der Deutschlernenden, aufgeteilt nach Vokal und Versuchsperson

Zusammenfassend zeigt sich, dass die deutschen Muttersprachler und Muttersprachlerinnen in der Identifikationsleistung insgesamt besser abschneiden als die Deutschlernenden. Ferner scheinen auch auf einem hohen Sprachniveau (überwiegend C1) Sprachlernende bei basalen phonologischen Aufgaben Schwierigkeiten zu haben. Es lässt sich jedoch auch festhalten, dass sich der Laut /o:/ den Ergebnissen beider Gruppen nach als universell leicht zu identifizierender Vokal abzeichnet. Auch /e:/ und /ɛ:/ können als universell weniger problematische Vokale identifiziert werden, da hier sowohl die Deutschsprachigen als auch die Deutschlernenden akzeptable Identifikationsleistungen aufweisen. Im Gegenzug bereitet die Identifizierung der Vokale /u:/ und /y:/ in beiden Untersuchungsgruppen größere Probleme. Für die Deutschlernenden zusätzlich problematisch war der Laut /ø:/. Im Hinblick auf die Varianz der Antworten pro Probandengruppe und Vokal weist die Gruppe der Deutschlernenden außer bei /o:/ eine hohe Varianz im Antwortverhalten auf. Besonders die Vokale /ø:/ und /y:/ sind hier auffällig. Im Gegensatz dazu zeigt die Gruppe der Deutschsprachigen nur bei /u:/ und /y:/ eine höhere Varianz. Dement-

sprechend kann festgehalten werden, dass sich die Heterogenität der Gruppe der Sprachlernenden in Bezug auf ihre Sprachbiografie in den Ergebnissen widerzuspiegeln scheint.

### 3.4.2 Fehlerart

Ein genauer Blick auf die falsch perzipierten Laute ermöglicht es, differenziertere Aussagen über die Identifikationsleistungen zu treffen. Bezüglich der Gruppe der monolingual deutschsprachigen Versuchspersonen zeigt sich, dass /u:/ in 21,62% der Fälle als /o:/ identifiziert wurde (Tabelle 4). Der umgekehrte Fall trat nicht ein. Die Langvokale /y:/ und /ø:/ hingegen wurden in beide Richtungen miteinander verwechselt. Dabei zeigt sich eine deutliche Asymmetrie zugunsten der Identifikation von /y:/ als /ø:/. Zu 31,35% identifizierten die Teilnehmerinnen und Teilnehmer /y:/ als /ø:/. Allerdings hörten nur 0,81% den Langvokal /ø:/ als /y:/.

Stimulus	/u:/		/o:/	/y:/	/ø:/	
Fehler	/o:/	/ø:/	/m/	/ø:/	/y:/	/o:/
%	21,62	0,27	0,27	31,35	0,81	0,27

Tabelle 4: Fehlerarten der monolingual Deutschsprachigen. Häufigkeit (in %), zu der ein Stimulus mit einem anderen Laut (Fehler) verwechselt wurde

Darüber hinaus identifizierten je eine Versuchsperson einmalig /u:/ als /ø:/ und /ø:/ als /o:/, was jeweils 0,27% der Fehler des entsprechenden Stimulus entspricht. Ebenfalls einmalig notierte eine Versuchsperson den Buchstaben <m> nach Präsentation von /o:/. Da es sich um den ersten Stimulus handelte, den die Person hörte, ist davon auszugehen, dass der Fehler durch Unaufmerksamkeit bedingt war. Die in diesem Abschnitt genannten Fehler wiederholten sich weder innerhalb des Fragebogens einer Versuchsperson noch bei anderen Teilnehmenden aus der Gruppe der Deutschsprachigen und sind demnach zu vernachlässigen.

Gegenüber der geringen Anzahl der Fehlervarianten der Muttersprachlerinnen und Muttersprachler des Deutschen zeichnet sich bei den Deutschlernenden ein komplexeres Bild ab (Tabelle 5). So wurde /u:/ zu 24,29% als

/o:/ identifiziert, was in etwa dem Wert der Monolingual Deutschsprachigen entspricht. Zusätzlich hörten die Deutschlernenden in 2,86% der Fälle /y:/ statt /u:/. In einem Fall (entspricht 1,43%) konnte /u:/ gänzlich nicht einem Laut der deutschen Sprache zugeordnet werden. Dass /o:/ als /u:/ identifiziert wurde, lässt sich nur ein einziges Mal beobachten. In beiden Gruppen scheint die Verwechslung von /u:/ und /o:/ somit weitestgehend unidirektional zu sein.

Stimulus	/u:/			/o:/	/y:/			/ø:/			/ɛ:/		/e:/	
Fehler	/o:/	/y:/	-	/u:/	/ø:/	/u:/	/y:/	/o:/	/u:/	/e:/	/a:/	/ɛ:/	/i:/	
%	24,29	2,86	1,43	1,43	31,43	5,71	22,86	7,14	8,57	22,86	1,43	12,85	5,71	

Tabelle 5: Fehlerarten der Deutschlernenden. Häufigkeit (in %), zu der ein Stimulus mit einem anderen Laut (Fehler) verwechselt wurde

Auch die durchschnittliche Häufigkeit, mit der /y:/ als /ø:/ notiert wurde, stimmt mit 31,43% mit den Ergebnissen der Deutschsprachigen überein. Die Deutschlernenden hörten /y:/ zusätzlich zu 5,71% als /u:/. Weniger eindeutig als in der Gruppe der deutschen Muttersprachlerinnen und Muttersprachler ist die Asymmetrie von /y:/ und /ø:/. 22,68% der präsentierten /ø:/ wurden von den Deutschlernenden als /y:/ identifiziert. Zudem notierten die Versuchspersonen dieser Gruppe in 7,14% der Fälle /o:/ und in 8,57% /u:/ statt /ø:/.

Des Weiteren schienen die Nicht-Muttersprachlerinnen und Nicht-Muttersprachler die Langvokale /ɛ:/ und /e:/ miteinander zu verwechseln. Auch hierbei herrscht Asymmetrie, da /ɛ:/ zu 22,86% als /e:/ und /e:/ zu 12,85% als /ɛ:/ identifiziert wurde. Der Vokal /ɛ:/ wurde zusätzlich in einem Fall (1,43%) als /a:/ notiert. Der Stimulus /e:/ wurden in 5,71% der Male als /i:/ gehört.

Insgesamt lässt sich erkennen, dass in erster Linie die hohen Vokale /u:/ und /y:/ Probleme bei der Identifikation bereiten. Diese wurden hauptsächlich mit den mittleren deutschen Vokalen verwechselt, die ihnen im Vokalraum benachbart sind, die also ähnliche F2-Werte und höhere F1-Werte aufweisen (Abb. Formantkarte in Kapitel 3.2). Hier findet sich auch die größte Varianz zwischen den Versuchspersonen der jeweiligen Gruppen. Im Gruppenver-

gleich fällt auf, dass die Deutschlernenden neben der prävalenten Fehlidentifikation von /o:/ statt /u:/ und /ø:/ statt /y:/ auch mehr Fehlervariationen aufweisen als die Deutschsprachigen.

## 4 Diskussion und Ausblick

Wie zu Beginn des Artikels in Kapitel 1 beschrieben, hat die Muttersprache aus psycho- und neurolinguistischer Perspektive einen entscheidenden Einfluss auf das Fremdsprachenlernen und damit auch auf die Leistungen der Lernenden. Jedoch wird bei Leistungsbeurteilungen von Sprachprüfungen der Einfluss der Muttersprache nicht explizit als Bewertungskriterium berücksichtigt (vgl. Eckes 2005). Die Tatsache, dass der Sprachhintergrund der Prüflinge nicht in die Bewertung mit einbezogen wird, liegt daran, dass nicht der Lernaufwand, sondern das Sprachniveau bewertet wird (vgl. Plassmann & Zeidler 2014: 239). Aus psycholinguistischer Perspektive ist dieses Vorgehen kritisch zu betrachten, da viele Studien in den letzten Jahrzehnten nachweisen konnten, dass die Muttersprache einen entscheidenden Einfluss auf das Fremdsprachenlernen und die Lauterkennung hat (vgl. z.B. Best & Tyler 2007; Flege 2002, 2003; Strange & Shafer 2008). Dabei wurde auch gezeigt, dass fremdsprachliche Laute, die einer L1-Lautkategorie ähnlich sind, schwieriger zu unterscheiden sind als Laute, die keine Ähnlichkeit mit einem muttersprachlichen Laut aufweisen (vgl. Flege 1995). Aus didaktischer Sicht scheint die Nichtberücksichtigung des Sprachhintergrundes der Prüflinge demzufolge insofern problematisch, als hier ein Ungleichgewicht in den Erfolgchancen bestehen kann. Durch die enge Verknüpfung von Sprachperzeption und -produktion (vgl. z.B. Flege et al. 1997) beim Fremdsprachenlernen schlägt sich diese Chancenungleichheit nicht nur in Hörverstehens- sondern auch in Produktionsaufgaben nieder. Eine Anpassung der Prüfungsvorbereitung und -konzeption hinsichtlich der Berücksichtigung des sprachlichen Hintergrundes könnte zu mehr Chancengleichheit und Fairness in Prüfungssituationen führen.

Deshalb hatte die vorliegende Studie zum Ziel, den Einfluss der sprachlichen Heterogenität auf die Erkennungsleistung (nicht-muttersprachlicher) deutscher Langvokale zu untersuchen. Hierzu wurde eine heterogene Gruppe



Deutschlernender sowie eine Gruppe Deutschsprechender untersucht. Die Deutschlernenden hatten in etwa das gleiche Sprachniveau, die Muttersprache der Teilnehmenden wurde zwar erfasst, diente jedoch nicht als Ausschlusskriterium. Daraus resultiert eine heterogene Gruppenzusammensetzung der Deutschlernenden mit sehr unterschiedlichen Herkunftssprachen. Aus diesem Grund lässt sich ein Vergleich zwischen den Ergebnissen der Studie und der realen Lehr- und Prüfungssituation ziehen. Aufgabe der Teilnehmenden war es, 60 Hörbeispiele sechs verschiedener standarddeutscher Langvokale richtig zu identifizieren. Hierbei wurden zwei Maße, Akkuratheit und Fehlerart, betrachtet und sowohl innerhalb der Teilnehmergruppen als auch im Gruppenvergleich analysiert. Übergeordnet lässt sich zunächst, wie vorhergesagt, feststellen, dass die Identifikationsleistung der Muttersprachlerinnen und Muttersprachler bei allen Vokalen über den Leistungen der Deutschlernenden liegt. Tiefgehend sollte die Studie jedoch zwei Forschungsfragen beantworten:

1. Gibt es Laute, die generell einfacher oder schwieriger zu erkennen sind?
2. Ist die Heterogenität der Sprachlernenden im Ergebnis der Identifikationsleistung erkennbar?

In Bezug auf die erste Forschungsfrage zeichnet sich ab, dass drei der untersuchten sechs Laute generell einfacher bzw. schwieriger zu benennen sind. Diese Befunde werden dadurch gestützt, dass die Ergebnisse beider Sprachgruppen, sowohl Deutschsprechende als auch Deutschlernende, ähnlich hohe oder niedrige Identifikationsleistungen bei bestimmten Lauten aufweisen. Beispielsweise scheint der Langvokal /o:/ leichter zu identifizieren zu sein als die anderen Vokale. In beiden Sprachgruppen lag die Akkuratheit bei fast 100 Prozent. Zusätzlich dazu scheinen die Langvokale /u:/ und /y:/ generell perzeptiv schwieriger zu identifizieren sein, da hier beide Sprachgruppen ähnlich niedrige Akkuratheitsraten erreichten. Die korrekte Erkennung des Langvokals /y:/ scheint die größte Herausforderung zu sein. Ungewöhnlich an diesem Befund ist, dass der mittlere Langvokal /o:/ leichter zu erkennen ist als die hohen Vokale /u:/ und /y:/. Im Gegensatz zu den vorliegenden Ergebnissen zeigte sich in behavioralen Studien mit Sprachlernenden, dass Randvokale

im Vokalraum aufgrund ihrer prägnanten akustisch-artikulatorischen Eigenschaften perzeptuell leichter zu verarbeiten sind (vgl. z.B. Polka & Bohn 2011). Warum bei der vorliegenden Studie die Erkennungsmuster umgekehrt sind, kann zu diesem Zeitpunkt nicht abschließend geklärt werden und bedarf daher weiterer Untersuchungen.

Bezüglich der zweiten Forschungsfrage kann konstatiert werden, dass die Heterogenität des Sprachhintergrundes bei den Sprachlernenden einen Einfluss auf die Identifikation deutscher Langvokale zu haben scheint. Mit Blick auf die Akkuratheit der Erkennung der Vokale /e:/ und /ɛ:/ zeigt sich, dass die Sprachlernenden bei beiden Vokalen unter gleichzeitiger hoher Varianz der individuellen Ergebnisse eine schlechtere Leistung aufweisen als die muttersprachlichen Teilnehmenden. Die höhere Varianz könnte hierbei für einen Einfluss des Sprachhintergrundes sprechen. Ein ähnliches, noch deutlicheres Bild zeichnet sich bei /ø:/ ab. Während auch hier die muttersprachlichen Probandinnen und Probanden über eine sehr hohe Identifikationsleistung verfügen, fällt die Leistung der nicht-muttersprachlichen Deutschlernenden deutlich ab. Zugleich zeigt sich hierbei erneut eine hohe individuelle Varianz. Insgesamt kann darüber hinaus festgehalten werden, dass die heterogene Gruppe der Deutschlernenden bei fünf der sechs getesteten Vokale eine deutliche höhere Zwischensubjektvarianz aufweisen als die deutschen Muttersprachlerinnen und Muttersprachler. Diese erhöhte Varianz der Leistungen spricht für einen Einfluss der jeweiligen Muttersprache auf die Identifikation nicht-nativer Vokale.

Auch die Analyse der Fehlerarten zeigt einen Einfluss der Heterogenität des sprachlichen Hintergrunds bei der Gruppe der Deutschlernenden auf die Identifikationsleistung. Während die Gruppe der monolingual Deutschsprechenden nur wenige und sehr begrenzte Fehlertypen zeigt, ist das Bild in der Gruppe der Deutschlernenden deutlich diverser. Insgesamt fällt für die Letztere auf, dass mehr und unterschiedliche Fehler gemacht werden. Darüber hinaus unterscheiden sich die Fehlertypen oftmals von denen der Deutschsprechenden. Diese Feststellungen treffen insbesondere für die Laute /u:/, /ɛ:/, /e:/, /y:/ und /ø:/ zu. Auch diese Diversität und die veränderten Fehlerarten im Vergleich zu denen aus der Gruppe der Deutschsprechenden lassen auf einen Einfluss der jeweiligen Muttersprache hinsichtlich der Erkennungsleistung schließen.

Zusammenfassend lässt sich festhalten, dass es sowohl Laute zu geben scheint, die für beide Sprachgruppen generell einfach oder schwieriger zu identifizieren sind, als auch, dass die Heterogenität des Sprachhintergrunds bei Deutschlernenden einen Einfluss auf die Ergebnisse der Erkennungsleistung hat.

Die vorliegende Untersuchung stellt eine psycholinguistische Pilotstudie dar und lässt daher an einigen Punkten Fragen offen. Wie bereits erwähnt, ist im Moment nicht zu erklären, weshalb Vokale der mittleren Zungenhöhe sowohl von Muttersprachlerinnen und Muttersprachlern als auch von Deutschlernenden leichter zu identifizieren waren als hohe Vokale. Durch die unterschiedliche Größe der beiden Sprachgruppen ergeben sich ebenfalls Limitationen. Gerade die Gruppe der Deutschlernenden war mit sieben Teilnehmenden verhältnismäßig klein. Daher würde sich eine weitere Untersuchung mit einer größeren Probandengruppe an Deutschlernenden empfehlen. Zusätzlich könnten auch erneute Testungen mit gezielt kontrollierten Muttersprachen sinnvoll sein, um den Einfluss von bestimmten Muttersprachen auf die Erkennung verschiedener deutscher Vokale noch besser nachvollziehen zu können. Strange & Shafer (2008) führten darüber hinaus an, dass eine Benennung mittels orthografischer Antwort nicht ausreichend repräsentativ ist, besonders dann, wenn die Teilnehmenden nicht über ein ausreichend hohes Sprachniveau verfügen. Obwohl in der vorliegenden Studie die Probandinnen und Probanden mindestens ein Sprachniveau von B2 hatten und daher angenommen werden kann, dass die Aufgabe den Teilnehmenden keine Probleme bereitete, könnte sich bei erneuten Untersuchungen eine Anpassung der Aufgabenstellung anbieten. Denkbar wäre hier beispielsweise eine Zuordnungsaufgabe, bei der Grapheme angezeigt werden, die dem jeweiligen Laut zugeordnet werden müssen. Zusätzlich dazu könnte eine Anpassung der Präsentation der Laute vorgenommen werden. Die isolierte Präsentationsweise ist für die Teilnehmenden eine ungewöhnliche und unnatürliche Perzeptionssituation. Wie Studien zeigen konnten, wird die Erkennung phonologischer Repräsentationen nicht nur mit zunehmender Größe des Vokabulars (vgl. Bundgaard-Nielsen et al. 2011) erleichtert, sie ist auch insgesamt kontextabhängig. Es wurde erwiesen, dass sich Vokale einfacher identifizieren lassen, sobald diese von Konsonanten umgeben sind (vgl. z.B. Strange et al. 1979; Rakerd 1984). Aus diesem Grund könnte sich

bei weiteren Studien eine syllabische Präsentationsweise der Laute (z.B. /tu:/) anbieten. Insgesamt wären auch Untersuchungen mittels Pseudowörtern oder Minimalpaaren interessant, um die Untersuchung zusätzlich in einer natürlicheren Perzeptionssituation durchzuführen.

Trotz der genannten Limitationen liefert die vorliegende Studie Evidenzen für den Einfluss der Heterogenität des sprachlichen Hintergrunds auf die phonologischen Erkennungsleistungen. Bedingt durch diese sprachliche Diversität werden demnach die verschiedenen Teilnehmerinnen und Teilnehmer vor unterschiedliche Herausforderungen im Aufgabenkontext gestellt. Gestützt durch psycho- und neurolinguistische Erkenntnisse (siehe Kapitel 1) ist daher jedoch auch anzunehmen, dass damit ebenfalls unterschiedliche Herausforderungen beim Fremdsprachenlernen einhergehen. Besonders auffällig an dieser Identifikationsstudie war, dass selbst monolinguale Deutschsprecherinnen und Deutschsprecher bei einigen Lauten erhebliche Schwierigkeiten mit der korrekten Erkennung der Laute hatten. Dies zeigte sich einmal durch niedrigere Akkuratheit, aber auch durch hohe individuelle Varianzen in den Erkennungsleistungen (z.B. /y:/). Wenn selbst muttersprachlich Deutschsprechende Schwierigkeiten bei dieser Aufgabe haben, so ist es nicht verwunderlich, dass Deutschlernende mit heterogenen sprachlichen Hintergründen hier besondere Schwierigkeiten und noch größere individuelle Varianz aufzeigen.

Aus diesem Grund könnte eine direkte Berücksichtigung der Herkunftssprachen bei der Bewertung von Sprachprüfungen sinnvoll sein. Durch Einbeziehung des Sprachhintergrunds in die Leistungsbeurteilung könnte eine noch größere Chancengleichheit bzw. Fairness der Sprachprüfungen erzielt werden. Da Sprachprüfungen jedoch das Sprachniveau unabhängig vom Sprachhintergrund bewerten sollen, wäre eine nachträgliche Korrektur der Prüfungsergebnisse für den jeweiligen Sprachhintergrund nicht zielführend. Vielmehr könnte es beispielsweise sinnvoll sein, Lauterkennungsaufgaben auch in den Prüfungskontext mit einzubeziehen. Wenn in Prüfungen diese immanenten Schwierigkeiten miterhoben werden, könnte bei Nichtbestehen der Hörverstehens- und Sprachproduktionsaufgaben anhand dieser Daten überprüft werden, ob dies nicht auf die Lernleistungen, sondern auf den Sprachhintergrund zurückzuführen ist. Denkbar wäre, den Prüfungsteilnehmenden eine Wiederholung des Hörverstehens, da in den Prüfungen auf höherem Niveau eine Wiederholung

der ganzen Prüfung unabhängig des vorherigen Ergebnisses die Regel ist, zu ermöglichen. So könnte in der Vorbereitung auf eine Prüfungswiederholung der Fokus auf die Lauterkennung und Aussprache gelegt werden, um die Chancen auf erfolgreiches Bestehen zu erhöhen.

Offensichtlich ist aber auch, dass die Berücksichtigung des Sprachhintergrundes selbst Herausforderungen mit sich bringt. Aus diesem Grund bietet sich aus psycho- und neurolinguistischer Sicht eine verstärkte Fokussierung auf gezielte Trainingseinheiten in der Lehre und in der Prüfungsvorbereitung an. Die Lauterkennung ist eine Grundfertigkeit, die darüber hinaus positiv mit der Lautproduktion korreliert. Während Böttger (2016) beim frühen Fremdsprachenlernen gezielte Lauterkennungsübungen vorschlägt, kann aus psycho- und neurolinguistischer Sicht argumentiert werden, dass diese basalen Fertigkeiten auch bei höherem Sprachniveau gezielt trainiert werden sollte. Wie zahlreiche Studien zeigten, ist nicht nur das korrekte Lernen von nicht-nativen Lauten im Erwachsenenalter möglich (vgl. Golestani & Zatorre 2009; Logan et al. 1991; MacKain et al. 1981), durch die gezielte Übung der perzeptiven Fähigkeiten wird gleichzeitig auch die korrekte Produktion gefördert (vgl. z.B. Perkell et al. 2004).

Zusätzlich konnten einige Studien gezielte Trainingseffekte sowohl auf behavioraler als auch auf neuronaler Ebene zeigen. Iverson et al. (2011) testeten in einer behavioralen Studie die Identifikations- und Diskriminationsfähigkeit vor und nach gezielten Identifikationstrainingseinheiten. Sie fanden eine signifikant bessere Identifikationsleistung nach Absolvierung gezielter Übungsperioden mit gleichzeitig geringen Verbesserungen in der Lautdiskrimination und -produktion. Ähnliche Befunde konnten in weiteren Studien mit unterschiedlichen Lauten aus verschiedenen Sprachen reproduziert werden (vgl. Golestani & Zatorre 2008; Shinohara & Iverson 2018). Auch neuronal wirkt sich gezieltes und wiederholtes Training positiv auf die Erkennung und Verarbeitung von Lauten aus. So kommt es durch wiederholte Übungen zu veränderten Aktivierungen im Gehirn. Golestani & Zatorre (2004) etwa konnten zeigen, dass ohne Übungen nur Areale nicht-nativer Lautverarbeitung aktiviert werden, während im Anschluss an Übungseinheiten nach und nach auch Areale muttersprachlicher Lautverarbeitung mit aktiviert werden. Auch diese Befunde stützen die Sichtweise, dass Übungen zur Verbesserung basaler

phonologischer Fertigkeiten auch bei Sprachkursen höheren Niveaus eine stärkere Rolle spielen sollten. Durch die reziproke Verknüpfung von Sprachperzeption und -produktion würde das Training dieser basalen phonologischen Fähigkeiten ebenso auf Ebene der Lautproduktion zum Tragen gekommen. Im Umkehrschluss bedeutet dies darüber hinaus auch, dass Aussprachetrainings und Lautproduktionsübungen auch in hohen Sprachniveaus zielführend für die Verbesserung von Hörverstehensleistungen sein können.

Abschließend kann zweierlei konstatiert werden: 1. Die vorliegenden Ergebnisse sprechen dafür, die phonologische Spracherfahrung der Lernenden in der Lehre und der Prüfungsvorbereitung noch intensiver zu berücksichtigen. 2. Es ist effektiv, psycho- und neurolinguistische Befunde noch stärker in die Prüfungsentwicklung mit einzubeziehen (ähnlich wie dies im Bereich Neurodidaktik bereits geschieht).

## Bibliografie

- ADAMCZAK-KRYSZTOFOWICZ, S. & LIMBACH, C. (2019): „Hören“, in: JEUK, S. & SETTINIERI, J. (Hrsg.): *Sprachdiagnostik Deutsch als Zweitsprache*, Berlin, Boston: De Gruyter: 387–412, abrufbar unter: <https://doi.org/10.1515/9783110418712-016>, 23.07.2020.
- AICHERT, I., MARQUARDT, C. & ZIEGLER, W. (2005): „Frequenzen sublexikalischer Einheiten des Deutschen: CELEX-basierte Datenbanken“, in: *Neurolinguistik 19/1–2*: 55–81.
- ARCHILA-SUERTE, P. et al. (2012): „Age of acquisition and proficiency in a second language independently influence the perception of non-native speech“, in: *Bilingualism: Language and Cognition 15/1*: 190–201, abrufbar unter: <https://doi.org/10.1017/S1366728911000125>, 23.07.2020.
- ARCHILA-SUERTE, P., ZEVI, J. & HERNANDEZ, A. E. (2015): „The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds“, in: *Brain and Language 141*: 35–49, abrufbar unter: <https://doi.org/10.1016/j.bandl.2014.11.005>, 23.07.2020.

- ARRAS, U., MÜLLER-KARABIL, A. & ZIMMERMANN, S. (2013): „On equal footing? Accommodations for disabled candidates in the TestDaF“, in: TSAGARI, D. & SPANOUDIS, D. (Hrsg.): *Assessing L2 students with learning and other disabilities*, Newcastle upon Tyne, UK: Cambridge Scholars: 271–286.
- BAAAYEN, R. H., PIEPENBROCK, R. & GULIKERS, L. (1995): *The CELEX Lexical Database* (CD-ROM), Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- BEHRMANN, M., GENG, J. J. & SHOMSTEIN, S. (2004): „Parietal cortex and attention“, in: *Current Opinion in Neurobiology* 14/2: 212–217, abrufbar unter: <https://doi.org/10.1016/j.conb.2004.03.012>, 23.07.2020.
- BEST, C. T., MCROBERTS, G. W. & GOODELL, E. (2001): „Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system“, in: *The Journal of the Acoustical Society of America* 109/2: 775–794, abrufbar unter: <https://doi.org/10.1121/1.1332378>, 23.07.2020.
- BEST, C. T., & TYLER, M. D. (2007): „Nonnative and second-language speech perception: Commonalities and complementarities“, in: BOHN, O.-S. & MUNRO, M. J. (Hrsg.): *Language Learning & Language Teaching* (Bd. 17), Amsterdam: J. Benjamins Publishing: 13–34, abrufbar unter: <https://doi.org/10.1075/llt.17.07bes>, 23.07.2020.
- BÖTTGER, H. (2016): *Neurodidaktik des frühen Sprachenlernens: Wo die Sprache zuhause ist*, Bad Heilbrunn: Verlag Julius Klinkhardt.
- BUNDEGAARD-NIELSEN, R. L., BEST, C. T. & TYLER, M. D. (2011): „Vocabulary Size is Associated with Second-Language Vowel Perception Performance in Adult Learners“, in: *Studies in Second Language Acquisition* 33/3: 433–461, abrufbar unter: <https://doi.org/10.1017/S0272263111000040>, 23.07.2020.
- CALLAN, D. E. et al. (2004): „Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models“, in: *NeuroImage* 22/3: 1182–1194, abrufbar unter: <https://doi.org/10.1016/j.neuroimage.2004.03.006>, 23.07.2020.

- CINDARK, I. et al. (2019): *Perspektive Beruf. Mündliche Kompetenz von Teilnehmenden an Integrationskursen und Vorschläge für die Praxis*, Mannheim: Leibniz-Institut für Deutsche Sprache.
- COSTA, A. & SEBASTIÁN-GALLÉS, N. (2014): „How does the bilingual experience sculpt the brain?“, in: *Nature Reviews Neuroscience* 15/5: 336–345, abrufbar unter: <https://doi.org/10.1038/nrn3709>, 23.07.2020.
- COUNCIL OF EUROPE (2018): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion Volume with New Descriptors*, abrufbar unter: <https://www.coe.int/en/web/common-european-framework-reference-languages>, 30.07.2020.
- DORMAN, M. F. et al. (1985): „Phonetic identification by elderly normal and hearing-impaired listeners“, in: *The Journal of the Acoustical Society of America* 77/2: 664–670, abrufbar unter: <https://doi.org/10.1121/1.391885>, 23.07.2020.
- ECKES, T. (2003): „Qualitätssicherung beim TestDaF: Konzepte, Methoden, Ergebnisse“, in: *Fremdsprachen und Hochschule* 69: 43–68.
- ECKES, T. (2005): „Evaluation von Beurteilungen: Psychometrische Qualitätssicherung mit dem Multifacetten-Rasch-Modell“, in: *Zeitschrift für Psychologie* 213/2: 77–96, abrufbar unter: <https://doi.org/10.1026/0044-3409.213.2.77>, 23.07.2020.
- ECKES, T. (2008): „Assuring the quality of TestDaF examinations: A psychometric modelling approach“, in: TAYLOR, L. & WEIR, C. J. (Hrsg.): *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity: Proceedings of the ALTE Berlin Conference, May 2005*, Cambridge: Cambridge University Press: 157–178.
- ECKES, T. (2013): „A study of differential item functioning in the TestDaF reading and listening sections“, in: GALACZI, E. D. & WEIR, C. J. (Hrsg.): *Exploring language frameworks: Proceedings of the ALTE Kraków Conference 07/2011*, Cambridge: Cambridge University Press: 362–388.
- ESCUADERO, P., SIMON, E. & MITTERER, H. (2012): „The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception“, in: *Journal of Phonetics* 40/2: 280–288, abrufbar unter: <https://doi.org/10.1016/j.wocn.2011.11.004>, 23.07.2020.



- FARIS, M. M. (2017): *Perceptual assimilation, discrimination, and acquisition of non-native and second-language vowels assimilated as uncategoryed* [Doctoral dissertation], Sydney: Western Sydney University.
- FLEGE, J. E. (1995): „Second Language Speech Learning Theory, Findings, and Problems“, in: STRANGE, W. (Hrsg.): *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium, MD: York Press: 233–277.
- FLEGE, J. E. (2002): „Interactions between the Native and Second-language Phonetic Systems“, in: BURMEISTER, P. et al. (Hrsg.): *An Integrated View of Language Development: Papers in Honor of Henning Wode*, Trier: Wissenschaftlicher Verlag: 217–244.
- FLEGE, J. E. (2003): „Assessing constraints on second-language segmental production and perception“, in: MEYER, A. & SCHILLER, N. (Hrsg.): *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, Berlin: Mouton de Gruyter: 319–355.
- FLEGE, J. E., BOHN, O.-S. & JANG, S. (1997): „Effects of experience on non-native speakers’ production and perception of English vowels“, in: *Journal of Phonetics* 25/4: 437–470, abrufbar unter: <https://doi.org/10.1006/jpho.1997.0052>, 23.07.2020.
- FLEGE, J. E. & MACKAY, I. R. A. (2004): „Perceiving Vowels in a Second Language“, in: *Studies in Second Language Acquisition* 26/1: 1–34, abrufbar unter: <https://doi.org/10.1017/S0272263104026117>, 23.07.2020.
- FRY, D. B. et al. (1962): „The Identification and Discrimination of Synthetic Vowels“, in: *Language and Speech* 5/4: 171–189, abrufbar unter: <https://doi.org/10.1177/002383096200500401>, 23.07.2020.
- GERRITS, E. & SCHOUTEN, M. E. H. (2004): „Categorical perception depends on the discrimination task“, in: *Perception & Psychophysics* 66/3: 363–376, abrufbar unter: <https://doi.org/10.3758/BF03194885>, 23.07.2020.
- GOLESTANI, N. et al. (2006): „Brain Structure Predicts the Learning of Foreign Speech Sounds“, in: *Cerebral Cortex* 17/3: 575–582, abrufbar unter: <https://doi.org/10.1093/cercor/bhk001>, 23.07.2020.

- GOLESTANI, N. & ZATORRE, R. J. (2004): „Learning new sounds of speech: Reallocation of neural substrates“, in: *NeuroImage* 21/2: 494–506, abrufbar unter: <https://doi.org/10.1016/j.neuroimage.2003.09.071>, 23.07.2020.
- GOLESTANI, N. & ZATORRE, R. J. (2009): „Individual differences in the acquisition of second language phonology“, in: *Brain and Language* 109/2–3: 55–67, abrufbar unter: <https://doi.org/10.1016/j.bandl.2008.01.005>, 23.07.2020.
- GREIN, M., NAGELS, A. & RIEDINGER, M. (2020): *Neurodidaktik aktuell*. Manuskript in Vorbereitung.
- GURUNANDAN, K., CARREIRAS, M. & PAZ-ALONSO, P. M. (2019): „Functional plasticity associated with language learning in adults“, in: *NeuroImage* 201, abrufbar unter: <https://doi.org/10.1016/j.neuroimage.2019.116040>, 23.07.2020.
- HÜNLICH, D. et al. (2018): *Wer besucht den Integrationskurs? Soziale und sprachliche Hintergründe von Geflüchteten und anderen Zugewanderten*, Mannheim: Institut für Deutsche Sprache.
- IVERSON, P., PINET, M. & EVANS, B. G. (2011): „Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels“, in: *Applied Psycholinguistics* 33/1: 145–160, abrufbar unter: <https://doi.org/10.1017/S0142716411000300>, 23.07.2020.
- JEONG, H. et al. (2007): „Cross-linguistic influence on brain activation during second language processing: An fMRI study“, in: *Bilingualism: Language and Cognition* 10/2: 175–187, abrufbar unter: <https://doi.org/10.1017/S1366728907002921>, 23.07.2020.
- KECKER, G. (2016): „Was macht eine gute Sprachprüfung aus? Qualitätssicherung beim TestDaF“, in: DRUMBL, J. & HORNING, A. (Hrsg.): *IDT 2013*, Bozen: Bozen-Bolzano University Press: 145–164.
- KIM, K. H. S. et al. (1997): „Distinct cortical areas associated with native and second languages“, in: *Nature* 388: 171–174, abrufbar unter: <https://doi.org/10.1038/40623>, 23.07.2020.
- KIM, S. Y. et al. (2016): „How does language distance between L1 and L2 affect the L2 brain network? An fMRI study of Korean–Chinese–English

- trilinguals“, in: *NeuroImage* 129: 25–39, abrufbar unter: <https://doi.org/10.1016/j.neuroimage.2015.11.068>, 23.07.2020.
- KUHL, P. K. (2004): „Early language acquisition: Cracking the speech code“, in: *Nature Reviews Neuroscience* 5/11: 831–843, abrufbar unter: <https://doi.org/10.1038/nrn1533>, 23.07.2020.
- KUHL, P. K. et al. (1992): „Linguistic experience alters phonetic perception in infants by 6 months of age“, in: *Science* 255/5044: 606–608, abrufbar unter: <https://doi.org/10.1126/science.1736364>, 23.07.2020.
- LI, P., LEGAULT, J. & LITCOFSKY, K. A. (2014): „Neuroplasticity as a function of second language learning: Anatomical changes in the human brain“, in: *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 58: 301–324, abrufbar unter: <https://doi.org/10.1016/j.cortex.2014.05.001>, 23.07.2020.
- LIU, H., & CAO, F. (2016): „L1 and L2 processing in the bilingual brain: A meta-analysis of neuroimaging studies“, in: *Brain and Language* 159: 60–73, abrufbar unter: <https://doi.org/10.1016/j.bandl.2016.05.013>, 23.07.2020.
- LOGAN, J. S., LIVELY, S. E. & PISONI, D. B. (1991): „Training Japanese listeners to identify English /r/ and /l/: A first report“, in: *The Journal of the Acoustical Society of America* 89/2: 874–886, abrufbar unter: <https://doi.org/10.1121/1.1894649>, 23.07.2020.
- MACKAIN, K. S., BEST, C. T. & STRANGE, W. (1981): „Categorical perception of English /r/ and /l/ by Japanese bilinguals“, in: *Applied Psycholinguistics* 2/4: 369–390, abrufbar unter: <https://doi.org/10.1017/S0142716400009796>, 23.07.2020.
- MARIAN, V., SPIVEY, M. & HIRSCH, J. (2003): „Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging“, in: *Brain and Language* 86/1: 70–82, abrufbar unter: [https://doi.org/10.1016/S0093-934X\(02\)00535-7](https://doi.org/10.1016/S0093-934X(02)00535-7), 23.07.2020.
- MARKS, D. (2015): „Prüfen sprachlicher Kompetenzen internationaler Studienanfänger an deutschen Hochschulen – Was leistet der TestDaF?“, in: *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 20/1: 21–39.
- MCALLISTER, R., FLEGE, J. E. & PISKE, T. (2002): „The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, Eng-

- lish and Estonian“, in: *Journal of Phonetics* 30/2: 229–258, abrufbar unter: <https://doi.org/10.1006/jpho.2002.0174>, 23.07.2020.
- NORRIS, D., MCQUEEN, J. M. & CUTLER, A. (2003): „Perceptual learning in speech“, in: *Cognitive Psychology* 47/2: 204–238, abrufbar unter: [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9), 23.07.2020.
- OHALA, D. K. (2008): „Phonological acquisition in a first language“, in: EDWARDS, J. G. H. & ZAMPINI, M. L. (Hrsg.): *Phonology and Second Language Acquisition*, Amsterdam, Philadelphia: J. Benjamins Publishing: 19–40.
- OSTERHOUT, L. et al. (2008): „Second-language learning and changes in the brain“, in: *Journal of Neurolinguistics* 21/6: 509–521, abrufbar unter: <https://doi.org/10.1016/j.jneuroling.2008.01.001>, 23.07.2020.
- PARK, H. R. P., BADZAKOVA-TRAJKOV, G. & WALDIE, K. E. (2012): „Language lateralisation in late proficient bilinguals: A lexical decision fMRI study“, in: *Neuropsychologia* 50/5: 688–695, abrufbar unter: <https://doi.org/10.1016/j.neuropsychologia.2012.01.005>, 23.07.2020.
- PARKER JONES, O. et al. (2012): „Where, When and Why Brain Activation Differs for Bilinguals and Monolinguals during Picture Naming and Reading Aloud“, in: *Cerebral Cortex* 22/4: 892–902, abrufbar unter: <https://doi.org/10.1093/cercor/bhr161>, 23.07.2020.
- PERANI, D. & ABUTALEBI, J. (2005): „The neural basis of first and second language processing“, in: *Current Opinion in Neurobiology* 15/2: 202–206, abrufbar unter: <https://doi.org/10.1016/j.conb.2005.03.007>, 23.07.2020.
- PERKELL, J. S. et al. (2004): „The distinctness of speakers’ productions of vowel contrasts is related to their discrimination of the contrasts“, in: *The Journal of the Acoustical Society of America* 116/4: 2338–2344, abrufbar unter: <https://doi.org/10.1121/1.1787524>, 23.07.2020.
- PLASSMANN, S. & ZEIDLER, B. (2014): „Taking decisions: Assessment for university entry“, in: *Language Learning in Higher Education* 4/1, abrufbar unter: <https://doi.org/10.1515/cercles-2014-0013>, 23.07.2020.
- POLKA, L. & BOHN, O.-S. (2003). „Asymmetries in vowel perception“, in: *Speech Communication* 41/1: 221–231, abrufbar unter: [https://doi.org/10.1016/S0167-6393\(02\)00105-X](https://doi.org/10.1016/S0167-6393(02)00105-X), 23.07.2020.

- POLKA, L. & BOHN, O.-S. (2011): „Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development“, in: *Journal of Phonetics* 39/4: 467–478, abrufbar unter: <https://doi.org/10.1016/j.wocn.2010.08.007>, 23.07.2020.
- RAKERD, B. (1984): „Vowels in consonantal context are perceived more linguistically than are isolated vowels: Evidence from an individual differences scaling study“, in: *Perception & Psychophysics* 35/2: 123–136, abrufbar unter: <https://doi.org/10.3758/BF03203892>, 23.07.2020.
- SCHOUTEN, M. E. H. & VAN HESSEN, A. J. (1992): „Modeling phoneme perception. I: Categorical perception“, in: *The Journal of the Acoustical Society of America* 92/4: 1841–1855, abrufbar unter: <https://doi.org/10.1121/1.403841>, 23.07.2020.
- SEBASTIAN, R., LAIRD, A. R. & KIRAN, S. (2011): „Meta-analysis of the neural representation of first language and second language“, in: *Applied Psycholinguistics* 32/4: 799–819, abrufbar unter: <https://doi.org/10.1017/S0142716411000075>, 23.07.2020.
- SEBASTIÁN-GALLÉS, N. ET AL. (2012): „Neuroanatomical markers of individual differences in native and non-native vowel perception“, in: *Journal of Neurolinguistics* 25/3: 150–162, abrufbar unter: <https://doi.org/10.1016/j.jneuroling.2011.11.001>, 23.07.2020.
- SENDLMEIER, W. F. & SEEBODE, J. (2006): *Formantkarten des deutschen Vokalsystems*, abrufbar unter: <https://www.kw.tu-berlin.de/menue/projekte/formantkarten>, 23.07.2020.
- SHINOHARA, Y. & IVERSON, P. (2018): „High variability identification and discrimination training for Japanese speakers learning English /r/-/l/“, in: *Journal of Phonetics* 66: 242–251, abrufbar unter: <https://doi.org/10.1016/j.wocn.2017.11.002>, 23.07.2020.
- STRANGE, W., EDMAN, T. R. & JENKINS, J. J. (1979): „Acoustic and Phonological Factors in Vowel Identification“, in: *Journal of Experimental Psychology: Human Perception and Performance* 5/4: 643–656.
- STRANGE, W. & SHAFER, V. L. (2008): „Speech perception in second language learners: The re-education of selective perception“, in: EDWARDS, J. G. H. & ZAMPINI, M. L. (Hrsg.): *Phonology and Second Language Acquisition*, Amsterdam, Philadelphia: J. Benjamins Publishing: 153–192.

- SULPIZIO, S. et al. (2020): „Bilingual language processing: A meta-analysis of functional neuroimaging studies“, in: *Neuroscience & Biobehavioral Reviews* 108: 834–853, abrufbar unter: <https://doi.org/10.1016/j.neubio-rev.2019.12.014>, 23.07.2020.
- WERKER, J. F. & TEES, R. C. (1984a): „Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life“, in: *Infant Behavior and Development* 7: 49–63.
- WERKER, J. F. & TEES, R. C. (1984b): „Phonemic and phonetic factors in adult cross-language speech perception“, in: *The Journal of the Acoustical Society of America* 75/6: 1866–1878, abrufbar unter: <https://doi.org/10.1121/1.390988>, 23.07.2020.
- YOKOYAMA, S. et al. (2013): „Cross-linguistic influence of first language writing systems on brain responses to second language word reading in late bilinguals“, in: *Brain and Behavior* 3/5: 525–531, abrufbar unter: <https://doi.org/10.1002/brb3.153>, 23.07.2020.

## Internetquellen

- BOERSMA, P. & WEENINK, D. (2019): *Praat*: Doing phonetics by computer (6.0.52) [Computer software], abrufbar unter: <http://www.praat.org>, 02.05.2019.



## Zu den Autorinnen und Autoren

**Christina Maria Ersch**, wissenschaftliche Mitarbeiterin des Masterstudiengangs Deutsch als Fremdsprache/Deutsch als Zweitsprache an der Universität Mainz, unterrichtet seit 2012 Deutsch als Fremdsprache an Universitäten und Sprachschulen. Sie ist lizenzierte Prüferin für die *telc* Prüfungen Deutsch B1–B2 und den Deutschtest für Zuwanderer. Ihre Forschungsinteressen liegen in der (Hochschul-)Didaktik, insbesondere in der Umsetzung der Kompetenzorientierung und Digitalisierung der Lehre, der interkulturellen Kommunikation und dem Einfluss von Emotionen im (Fremdsprachen)-Unterricht.

**Ann-Katrin Fierus** lehrte nach ihren Abschlüssen in Deutsch als Fremdsprache/Deutsch als Zweitsprache (M.A.) und Deutsch/Englisch (M.Ed.) an der Universität Mainz zunächst in einer Sprachschule in Arequipa, Peru. Hier stand neben der Sprachvermittlung vor allem der kulturelle Austausch im Vordergrund. Ihr Interessen- und Forschungsschwerpunkt liegt im Gebiet der interkulturellen Kommunikation. Nach ihrer Arbeit als Betreuerin und Übersetzerin für deutsche Passagiere auf Kreuzfahrtschiffen, unterrichtet sie nun Deutsch als Fremdsprache in Darmstadt.

**Marion Grein**, Leiterin des Masterstudiengangs Deutsch als Fremdsprache/Deutsch als Zweitsprache an der Universität Mainz, hat u.a. im Bereich Sprachwissenschaft und Schul- und Erwachsenenpädagogik promoviert und sich für Allgemeine Sprachwissenschaft, Vergleichende Sprachwissenschaft und Deutsch als Fremdsprache habilitiert. Sie ist national und international auf zahlreichen Konferenzen unterwegs. Ihre Schwerpunkte sind die Sprachlehrforschung, digitale Medien und interkulturelle Kompetenz. Sie ist u.a. Mitglied des Beirats Sprachen des Goethe-Instituts und im Vorstand der International Association for Dialogue Analysis.



**Nina Jehle** studierte Deutsch als Fremdsprache an der Johannes Gutenberg-Universität Mainz und ist Lehrkraft für fachspezifische Deutschkurse für internationale Studierende in der Universitätsmedizin. Neben Lehrtätigkeiten in Deutschland war sie mehrere Jahre als Dozentin für DaF an der Universität in der Türkei tätig. Derzeit leitet sie ‚Mediziner-Kurse‘ im Rahmen des DeuFöV-Spezialmoduls „Akademische Heilberufe“ und ist als Lehrbeauftragte im Bereich DaF/DaZ aktiv. Des Weiteren ist sie lizenzierte *telc*-Prüferin und nimmt u.a. regelmäßig Deutschtests für Zuwanderer ab. Darüber hinaus ist sie Online-Tutorin und Schulungsreferentin im Bereich Digitales Lernen.

**Miriam Riedinger**, wissenschaftliche Mitarbeiterin an der Johannes Gutenberg-Universität Mainz, forscht seit 2017 im Neurolinguistics Lab der Allgemeinen Sprachwissenschaft. Ihr Forschungsschwerpunkt liegt dabei in der Erkundung der neuronalen Prozesse der Lautperzeption, welche sie vorrangig durch elektrophysiologische Messungen (EEG) untersucht. Neben der Forschungstätigkeit ist sie mit der Lehre verschiedener psycho- und neurolinguistischer Seminare im Bachelor- und Masterstudiengang der Linguistik betraut.

**Virginia Sánchez Anguix** ist seit ihrem Studium der Angewandten Sprach- und Kulturwissenschaften (FASK in Germersheim) an der Universität Mainz Übersetzerin für Deutsch und Spanisch und seit vielen Jahren in der freien Wirtschaft tätig u.a. als Trainerin. Aktuell studiert sie Deutsch als Fremdsprache/Deutsch als Zweitsprache an der Johannes Gutenberg-Universität in Mainz, ist Tutorin für Grammatik und lehrt zur Zeit Deutsch als Fremdsprache im Auftrag der Frankfurter University of Applied Sciences.

**Joshua Ziegler** studiert Allgemeine und Vergleichende Sprachwissenschaft an der Johannes Gutenberg-Universität Mainz. Er ist zudem im Neurolinguistics Lab der JGU Mainz als wissenschaftliche Hilfskraft tätig. Neben der Phonetik und Phonologie beschäftigt er sich mit der Frage, wie neue und relevante Informationen sprachlich hervorgehoben und verarbeitet werden.

# DAF / DAZ IN FORSCHUNG UND LEHRE

- Bd. 1 Christina Maria Ersch (Hg.): Kompetenzen in DaF/DaZ.  
200 Seiten. ISBN 978-3-7329-0616-1
- Bd. 2 Christina Maria Ersch (Hg.): Evaluieren und Prüfen in DaF/DaZ.  
314 Seiten. ISBN 978-3-7329-0728-1
- Bd. 3 Kai Witzlack-Makarevich/Nadja Wulff/Coretta Storz (Hg.):  
Sushi, Sandmann, Sozialismus: Kultur- und Landeskunde der DDR.  
Ein Text- und Übungsbuch für den DaF-Unterricht.  
434 Seiten. ISBN 978-3-7329-0686-4

