

Beitrag Themenheft „Testen und Prüfen“

Almut Schön*

DaF und die Fachsprachenprüfungen für Ärzte, Apotheker und Pflegekräfte

German as a foreign language and the specialized language exams for health professionals

<https://doi.org/10.1515/infodaf-2022-0069>

Zusammenfassung: Erst seit wenigen Jahren müssen aus dem Ausland kommende Angehörige verschiedener Heilberufe, die in Deutschland eine Berufserlaubnis erlangen wollen, neben entsprechenden Fachkenntnissen auch Deutschkenntnisse in einem speziellen Prüfungsformat nachweisen. Im Folgenden werden die Fachsprachenprüfungen für Ärzte, Apotheker und Pflegekräfte¹ vorgestellt. Es wird diskutiert, inwiefern diese Prüfungen die Hauptgütekriterien von Tests erfüllen und welche Forschungsaufgaben sich hier für die angewandte DaF-Forschung ergeben.

Schlüsselwörter: Fachsprachenprüfungen, Testforschung, Gesprächsanalyse

Abstract: Only in the last few years health professionals who come from abroad and wish to obtain a professional license in Germany have been required to prove their knowledge of German in a specialized language exam in addition to the assessment of their professional knowledge. In the following, the specialized language exams for physicians, pharmacists and nurses are presented. It will be discussed to what extent these exams fulfil the main quality criteria of tests and which research tasks arise here for applied research in the field of German as a foreign language.

Keywords: specialized language exams, testing research, conversation analysis

¹ In diesem Text wird der Einheitlichkeit halber nur von Fachsprachenprüfungen gesprochen, obwohl sich oft auch der Terminus „Fachsprachprüfung“ findet. Sofern möglich, werden generische

*Kontaktperson: Almut Schön, E-Mail: almut.schoen@tu-berlin.de

1 Einleitung

Besonders in den ländlichen Regionen Deutschlands ist seit vielen Jahren ein Fachkräftemangel in verschiedenen Heilberufen zu beklagen. Ärzte und Pflegekräfte zählen zu den sogenannten Mangelberufen, aber auch in den Apotheken gibt es zu viele unbesetzte Stellen (vgl. Nier 2017; Bundesagentur für Arbeit 2021: 14). In diesen Branchen existiert daher seit Jahren neben Ausbildungsoffensiven auch eine Zuwanderung von qualifizierten Fachkräften aus dem Ausland. So hat sich beispielsweise die Zahl der in Deutschland berufstätigen Ärzte aus dem Ausland im Zeitraum von 1996 bis 2020 mehr als verfünffacht; zurzeit beträgt deren Anteil an der Gesamtärztezahl in Deutschland 13,7 Prozent (vgl. Bundesärztekammer 2020). Auch unter den in Deutschland berufstätigen Pflegekräften sind 13 Prozent aus dem Ausland (Radtko 2021; Nier 2017). Vor allem um die Patientensicherheit zu gewährleisten, hat die Gesundheitsministerkonferenz angesichts dieser Situation 2014 und 2019 beschlossen, dass Angehörige von Heilberufen nicht nur ihre beruflichen Kenntnisse, sondern auch Deutschkenntnisse auf bestimmten Niveaus des Gemeinsamen Europäischen Referenzrahmens für Sprachen (Trim et al. 2001) in einem definierten Prüfungsformat nachweisen müssen (Gesundheitsministerkonferenz 2014 und 2019). Angehörige akademischer Heilberufe, wie unter anderen Ärzte und Apotheker, müssen in einer besonderen Fachsprachenprüfung Kenntnisse auf dem Niveau C1 des GER dokumentieren, Pflegekräfte die Niveaustufe B2 im Deutschen als Fremdsprache erreichen. Die Fachsprachenprüfungen für Ärzte sowie für Apotheker führen in den meisten Bundesländern die Ärzte- bzw. die Apothekerkammern des Landes durch. Nur in drei von 16 Bundesländern werden für Ärzte andere Formate, etwa der TELC GmbH, anerkannt (Marburger Bund 2020). Fachsprachenprüfungen für Apotheker werden ausschließlich durch die Landesapothekerkammern abgenommen (ABDA 2021a). Welche Institutionen die Fachsprachenprüfungen für Pflegekräfte in Zukunft abnehmen werden, ist zurzeit noch nicht entschieden.

Die Gesundheitsministerkonferenz hat ein praxisorientiertes und vorwiegend mündliches Format für die Fachsprachenprüfungen für Heilberufe festgelegt, das aus den folgenden drei Elementen besteht (Gesundheitsministerkonferenz 2014 und 2019):

Personenbezeichnungen verwandt. In der Regel wird aufgrund des offiziellen Sprachgebrauchs der involvierten Institutionen sowie zur Vermeidung monotoner Textpassagen oder sprachlich heikler Sonderformen das generische Maskulinum benutzt.

- Gespräch mit einem Patienten (20 Minuten)
- Gespräch mit einem Angehörigen eines Heilberufs (20 Minuten), zum Beispiel mit einem Kollegen, der Vorgesetzten, einem Pfleger oder Arzt
- schriftliche Dokumentation (20 Minuten)

Angesichts der oben genannten Zahlen von Zuwanderern in diese Berufe und der gesamtgesellschaftlichen Bedeutung guter gesundheitlicher Versorgung ist klar, dass die Fachsprachenprüfungen für Heilberufe von kaum zu unterschätzender Praxisrelevanz sind und sich hier für Deutsch als fremde Fachsprache ein wichtiges Forschungs- und Wirkungsfeld auftut. Durch meine Mitwirkung an drei Evaluationsprojekten zu den Fachsprachenprüfungen für Apotheker, Ärzte und Pfleger im Zeitraum von 2015 bis 2021 konnten zu allen drei Prüfungen wichtige Daten erhoben werden, die im Folgenden systematisch ausgewertet werden sollen.² Im Zentrum dieses Artikels stehen dabei folgende Forschungsfragen:

- Inwiefern erfüllen die Fachsprachenprüfungen für Angehörige der Heilberufe die Hauptgütekriterien für Tests Objektivität, Reliabilität und Validität?
- Welche Methoden sind geeignet, um die Fachsprachenprüfungen wissenschaftlich zu untersuchen?

Nach einem einführenden Überblick über die erhobenen Daten sollen relevante Forschungspositionen und der aktuelle Forschungsstand besonders zu mündlichen Prüfungen referiert werden. Im Anschluss werden die Hauptgütekriterien in Bezug auf die Fachsprachenprüfungen näher analysiert. In einem Ausblick sollen die Fachsprachenprüfungen schließlich als diskursive Praktiken untersucht werden.

2 An dieser Stelle sei den Kooperationspartnern in der ABDA e.V., den Landesapothekerkammern, der Bayerischen Landesärztekammer sowie der passage gGmbH Hamburg ganz herzlich für die spannende und fruchtbare Zusammenarbeit und den stets offenen Austausch über die Fachgrenzen hinweg gedankt. Besonderer Dank gilt Frau Dr. Eckert-Lill, Frau Vogt, Frau Dr. Winter (ABDA e.V.), Herrn Dr. Burger und Herrn Reckzeh (BLAEK), Frau Beckmann-Schulz (passage gGmbH) sowie vielen weiteren Ansprechpartnerinnen und Ansprechpartnern. Ebenso gilt herzlicher Dank allen Ärzten, Apothekern und Pflegekräften, die damit einverstanden waren, dass in ihren Fachsprachenprüfungen hospitiert, getestet und zum Teil auch aufgenommen wurde. Ohne ihre große Unterstützung wären diese Forschungsprojekte nicht möglich gewesen. Weiterer Dank geht an Prof. Dr. Roelcke sowie die Kollegen und Kolleginnen vom Fachgebiet Deutsch als Fach- und Fremdsprache an der TU Berlin sowie an Frau Dr. Grunzig und Kerstin Zimmermann.

2 Überblick über die Forschungsdaten zu den Fachsprachenprüfungen für Heilberufe

2.1 Zu den Fachsprachenprüfungen für Apotheker

Von 2015 bis 2019 hatte ich die Gelegenheit, mit der ABDA – Bundesvereinigung Deutscher Apothekerverbände e.V. – zu den Fachsprachenprüfungen beratend zusammenzuarbeiten. Im Zentrum stand die Entwicklung und Fortschreibung eines Leitfadens für die Prüfungen, der unter anderem die Bewertungskriterien und den Bewertungsbogen umfasst. Dieser wurde mehrfach evaluiert und es bestand in diesem Zusammenhang auch die Möglichkeit, bei insgesamt 17 Fachsprachenprüfungen für Apotheker zu hospitieren. Die Ergebnisse dieser Evaluationen wurden in Prüferschulungen vorgestellt, auf nationaler Ebene diskutiert und die Ergebnisse dieser Diskussionen flossen in neue Fassungen des Leitfadens ein.

Da sich bei diesen High-Stakes-Prüfungen der Zugang zum Feld als schwierig erwies und viele Daten nicht zugänglich waren, war schnell klar, dass quantitative Forschungsansätze, wie sie in der Testforschung weit verbreitet sind, aufgrund der geringen Datenmenge nicht anwendbar sein würden. Zur Anwendung kamen daher qualitative Zugänge wie Beobachtungsprotokolle, Feldnotizen und vor allem Tonaufnahmen bei den Prüfungsgesprächen. In den beobachteten Prüfungen wurde folgendes Geschehen simuliert: Ein Patient kommt mit einem Rezept in die Apotheke und wird zur Abgabe und Anwendung des Medikaments sowie zu weiteren pharmazeutischen Aspekten durch den Prüfling beraten. Ein Beispiel aus einem Transkript soll diese Aufgabe illustrieren:

Tabelle 1: Transkriptausschnitt, Fachsprachenprüfung für Apothekerinnen und Apotheker, Kandidat 1, 2018

04	P	guten morgen #00:00:26-7#
05	Ap	<i>guten morgen, was kann isch für sie tun #00:00:30-4#</i>
06	P	ich hab n rezept dabei. #00:00:31-3#
07	Ap	<i>hm_hm ä fur sie selb? #00:00:34-4#</i>
08	P	ja. #00:00:35-2#
09	Ap	<i>hm_hm. äm der arzt hat ihnen beclometason spray verordnet #00:00:42-8#</i>
10	P	ja #00:00:42-8#
11	Ap	<i>ä wissen sie schon das medikament? haben sie erfahrung? #00:00:47-3#</i>

An diese Einleitung schließt sich dann die Aufgabe für den Apotheker an, zum Medikament und zur Einnahme sowie anderen pharmazeutischen Aspekten zu beraten. In einigen Prüfungen musste der Prüfling zusätzlich einen kurzen Monolog zu allgemeinen Fragen des Berufsalltags in der Apotheke halten, in anderen Prüfungen wurden weitere Gesprächsanlässe eingeführt, wie zum Beispiel die Bitte um ein verschreibungsfreies Medikament gegen alltägliche Beschwerden wie Kopfschmerzen oder Ähnliches durch den Patienten.

Der Patient berichtet im weiteren Fortgang des Gesprächs von einer unerwünschten Nebenwirkung des Medikaments, die schriftlich dokumentiert werden muss – dies ist der schriftliche Prüfungsteil. Auch hier gab es in einigen Prüfungen weitere Aufgaben, wie zum Beispiel das Verfassen einer berufstypischen Mail. Im zweiten mündlichen Prüfungsteil berichtet der Prüfling einem Apotheker über die Begegnung mit dem Patienten. Sowohl die Rolle des Kollegen als auch die des Patienten wird in den Prüfungen von unterschiedlichen Mitgliedern der Prüfungskommission übernommen.³

Die Prüfungen, bei denen ich im Rahmen der Zusammenarbeit mit der ABDA hospitieren konnte, fanden in Münster, Mainz, Stuttgart und München in den jeweiligen Landesapothekerkammern statt. Es gab entsprechend vier verschiedene Prüferteams, die zum Teil nur aus Apothekern, zum Teil aus Apothekern und Germanisten, Linguisten oder DaF-Didaktikern bestanden. Von 17 Prüflingen haben sich 15 mit Tonaufnahmen schriftlich einverstanden erklärt. Die Prüflinge kamen aus sechs Ländern: Syrien (8), Bosnien-Herzegowina (4), Polen (2), Spanien (1), Ungarn (1), Bulgarien (1). Diese Verteilung spiegelt sich auch in der Gesamtheit der Fachsprachenprüfungen für Apotheker 2020 wider: Von ca. 800 Prüfungskandidaten kamen mehr als 300 aus Syrien (ABDA 2021b: 83). Die Fachsprachenprüfung für Apotheker bestanden 14 der beobachteten 17 Prüflinge. Bei acht von 17 Prüfungen war es vorab möglich, mit den Prüflingen einen C-Test in Deutsch durchzuführen, dessen Ergebnisse zuvor mit den Stufen des GER kalibriert wurden (Schön/Johnson/Zimmermann 2014). Die Ergebnisse der C-Tests deuten auf Deutschkenntnisse von B1 bis C1. Aufgrund der geringen Probandenzahl und der wenigen Daten, die zur Verfügung stehen, können jedoch keine Aussagen zum Zusammenhang von C-Tests und den Ergebnissen der Fachsprachenprüfungen gemacht werden.

3 Aufgaben und Ablauf der Fachsprachenprüfungen werden auf den Seiten der zuständigen Landesapothekerkammern dargestellt; beispielhaft sei hier auf die ausführlichen Informationen der Apothekerkammer Nordrhein verwiesen: <https://www.aknr.de/download/apotheker/fachsprachenprfung.pdf> (letzter Abruf: 14.07.2022).

2.2 Zu den Fachsprachenprüfungen für Ärztinnen und Ärzte

Die Fachsprachenprüfungen für Ärzte waren die ersten Prüfungen, die speziell für Heilberufe entwickelt wurden. Sie wurden von den Ärztekammern entwickelt und werden in den allermeisten Bundesländern auch ausschließlich von den Kammern abgenommen. In den genannten Fachsprachenprüfungen wurden zuerst Prinzipien der Szenariendidaktik (Piepho 2003) angewandt, die sich dann auch in allen weiteren Fachsprachenprüfungen finden. Die Prüfung stellt ein möglichst realistisches Szenario aus dem Berufsleben ins Zentrum; geprüft werden komplexe Sprachhandlungen, keine isolierten Phänomene auf einzelnen sprachlichen Ebenen. Für Ärzte ist dieses Szenario ein Anamnesegespräch mit einem neuen Patienten. Im schriftlichen Teil wird das Gespräch in einem Arztbrief verschriftlicht, im dritten Teil wird der Fall mit einem Kollegen bzw. einer Kollegin besprochen, wobei Diagnosen und therapeutische Konzepte erörtert werden. Beide Prüfungsgespräche werden mit einem der Prüfer geführt, der dann in die Rolle des Patienten bzw. Kollegen schlüpft. Die Prüfungskommission besteht in der Regel aus drei Personen.⁴

Im Rahmen des Evaluationsprojekts in München konnten folgende Daten erhoben werden⁵:

- Audioaufnahmen von den mündlichen Prüfungsgesprächen (n = 14)
- C-Tests (n = 34)
- Bewertungsbögen für die Fachsprachenprüfungen für Ärzte (n = 33)
- Beobachtungen und Feldnotizen bei FSP (n = 6)
- Angaben zur L1, zur Deutschlerndauer, zum Abschlussjahr des Medizinstudiums, zur Facharztspezialisierung, zur Selbsteinschätzung des Deutschniveaus und zu den eigenen Erwartungen an die Fachsprachenprüfung (n = 34)

Ebenso wie bei den Fachsprachenprüfungen für Apotheker waren verschiedene Prüferkommissionen tätig und auch hier wurden verschiedene Szenarien durchgespielt. Alle Prüferkommissionen bestanden jeweils aus einem Germanisten respektive DaF-Didaktiker und zwei Ärzten. Die Prüflinge (N = 40) hatten 18 ver-

⁴ Auch hier kann vertiefend auf die Internetauftritte der Ärztekammern mit weiterführenden Informationen zu den Fachsprachenprüfungen verwiesen werden, beispielsweise auf den der Ärztekammer Berlin: <https://www.aekb.de/aerzt-innen/aus-dem-ausland-ins-ausland/fachsprachpruefung>

⁵ Von 2020 bis 2021 habe ich ein Projekt zur wissenschaftlichen Evaluation der Fachsprachenprüfungen für Ärztinnen und Ärzte in Bayern geleitet und hatte daher die Möglichkeit, in der Bayerischen Landesärztekammer Daten zu deren Fachsprachenprüfungen zu erheben.

schiedene L1. Die am häufigsten vertretenen L1 waren Russisch (9), Arabisch (8) und Spanisch (4), aber an der Studie nahmen auch Ärzte aus Afrika, Südamerika, China und verschiedenen zentralasiatischen Staaten sowie vielen europäischen Ländern teil.

Obwohl sich alle Prüflinge auf dem Niveau B2 respektive C1 einschätzten, zeigten die C-Tests, die hier in größerer Zahl durchgeführt werden konnten, ein weites Spektrum von A2 bis C1, was die Deutschkompetenzen betrifft. Anders als bei den Fachsprachenprüfungen für Apotheker konnten hier mehr und vielfältigere Daten erhoben werden, was zumindest einige einfache quantitative Verfahren der Testforschung ermöglicht.

2.3 Zu den Fachsprachenprüfungen für Pflegekräfte

Erst 2019 fasste die Gesundheitsministerkonferenz den Beschluss, dass auch Pflegekräfte neben ihren Fachkenntnissen Sprachkenntnisse in Deutsch auf dem Niveau B2 in einem separaten Sprachtest nachweisen müssen, ebenso wie Angehörige vieler weiterer Gesundheitsfachberufe. Die fünf Bundesländer Mecklenburg-Vorpommern, Niedersachsen, Schleswig-Holstein, Hamburg und Bremen betreiben gemeinsam das Norddeutsche Zentrum zur Weiterentwicklung in der Pflege (NDZ; siehe <https://www.pflege-ndz.de/startseite.html>). Das NDZ hat 2020 einen Hamburger Bildungsträger mit der Entwicklung einer neuen Fachsprachenprüfung für Pflege beauftragt – die Autorin war in dieses Projekt beratend eingebunden. Diese Fachsprachenprüfung hat denselben Aufbau wie die oben beschriebenen Fachsprachenprüfungen für Apotheker respektive Ärzte. Sie besteht aus einem Gespräch mit einem Patienten, das in geeigneter Form schriftlich dokumentiert werden muss, und einem abschließenden Gespräch mit einem Kollegen oder einer Kollegin, zum Beispiel bei der Dienstübergabe, wo noch einmal mündlich der jeweilige Fall dargestellt und das mögliche weitere pflegerische Vorgehen abgestimmt wird. Jeder Prüfungsteil umfasst 20 Minuten und insgesamt folgt auch diese Prüfung der Szenariendidaktik.

Im Rahmen des Projekts zur Prüfungsentwicklung wurden die entwickelten Szenarien auch in Prüfungen pilotiert. Dabei konnten diese Daten erhoben werden:

- Audioaufnahmen von den mündlichen Prüfungsteilen (n = 4)
- C-Test (n = 27) (siehe auch Abschnitt 2.1)
- Bewertungsbögen (n = 27)
- Beobachtungen und Feldnotizen bei Fachsprachenprüfungen (n = 4)
- Angaben zum Herkunftsland (n = 27)

Die Prüfungen fanden in Hamburg und Bremen mit zwei verschiedenen Prüferkommissionen statt, die jeweils aus einer DaF-Didaktikerin und einer Pflegefachfrau bestanden. Die Gespräche mit der Patientin respektive der Kollegin wurden jeweils mit einer Prüferin geführt, die für den jeweiligen Prüfungsteil die Rolle übernahm. Es wurden insgesamt zehn verschiedene Szenarien erprobt. Die Prüflinge stammten aus 15 verschiedenen Ländern, die größte Gruppe kam aus Brasilien ($n = 5$).⁶

Auch in diesem Projekt konnten C-Tests als Außenkriterium eingesetzt werden – die Ergebnisse verwiesen wiederum auf ein weites Spektrum von Deutschkenntnissen von A1 bis B2. Ebenso wie in den oben dargestellten Projekten sind auch hier einfache quantitative Analysen möglich und sollen unten dargestellt werden.

Zusammengefasst konnten in den drei geschilderten Evaluationsprojekten zu den Fachsprachenprüfungen die folgenden Daten erhoben werden:

Tabelle 2: Daten aus Evaluationsprojekten

Fachsprachenprüfung	N	Tonaufnahmen von mündlichen Prüfungsteilen	Bewertungsbögen	C-Tests
für Apotheker	17	15	–	8
für Ärzte	40	14	33	34
für Pflegekräfte	27	4	27	27

Nach einem kurzen Abriss über Forschungspositionen zu Testgütekriterien und zu mündlichen Tests sollen diese Daten im Folgenden qualitativ und quantitativ im Hinblick auf die Güte der Prüfungen ausgewertet werden.

3 Forschungsstand zu Prüfungen

3.1 Hauptgütekriterien in der klassischen Testtheorie

Im Kontext klassischer Testtheorie müssen Tests eine ganze Reihe von Gütekriterien erfüllen. Die Bedeutung und die Definition einzelner Kriterien sind umstritten, aber die Hauptgütekriterien Objektivität, Reliabilität und Validität gelten all-

⁶ Weitere Pflegekräfte kamen aus Bosnien, Indien, Jordanien, Kasachstan, Kolumbien, Kroatien, Mexiko, Nordmazedonien, Polen, Rumänien, Russland, Serbien, Spanien und Venezuela.

gemein als zentrale Anforderungen, denen gute Tests genügen müssen (Lienert/Raatz 1998; Moosbrugger/Kelava 2007; für die spezifische Diskussion von Sprachprüfungen Bachman/Palmer 1996; Grotjahn/Kleppin 2015; Fulcher/Davidson 2016). Diese drei Hauptgütekriterien sollen kurz näher dargestellt werden.

Objektivität bezeichnet die Unabhängigkeit der Testergebnisse vom Prüfer (Lienert/Raatz 1998: 7); es wird dabei zwischen Durchführungs-, Auswertungs- und Interpretationsobjektivität unterschieden. Eine hohe Durchführungsobjektivität wird zumeist über eine hohe Standardisierung der Testdurchführung und über Prüferschulungen erzielt. Eine hohe Auswertungsobjektivität ist zum Beispiel bei schriftlichen Prüfungen gegeben, wenn alle Antwortmöglichkeiten vorgegeben sind und ebenfalls klar ist, dass die richtige Antwort angekreuzt werden muss, zum Beispiel in einer Multiple-Choice-Aufgabe. Die Auswertung wird in diesem Fall unabhängig von den Auswertenden immer zum selben Ergebnis führen. Eine hohe Interpretationsobjektivität liegt dann vor, wenn die Prüfungsauswertung stets zur selben Gesamtwertung führt. Da in den Fachsprachenprüfungen nur die Bewertung *bestanden* oder *nicht bestanden* möglich ist, die anhand des Bewertungsbogens aufgrund von festen Regeln abgeleitet wird, sind die Ergebnisse interpretationsobjektiv; jedoch sollen Durchführungs- und Auswertungsobjektivität näher untersucht werden.

Die Reliabilität bezeichnet die Messgenauigkeit oder die Zuverlässigkeit der Messung (Lienert/Raatz 1998: 9). Sie kann durch den Abgleich von Ergebnissen in Paralleltests oder in Testwiederholungen bestimmt werden, aber auch durch die mathematische Berechnung der internen Konsistenz, wenn es organisatorisch nicht möglich ist, Tests zu duplizieren oder zu wiederholen. Die interne Konsistenz wird mit dem Koeffizienten Cronbachs α bestimmt, der im Prinzip die Korrelation der Items untereinander misst. Die Items wären in der Fachsprachenprüfung die verschiedenen Bewertungsskalen im Bewertungsbogen, die natürlich im Grundsatz dasselbe messen sollten, nämlich die Sprachkompetenz. Diese Berechnungen können aufgrund fehlender Daten für die Fachsprachenprüfung für Apotheker nicht durchgeführt werden; für die anderen beiden Fachsprachenprüfungen lässt sich eine Reliabilität mit gewissen Einschränkungen bestimmen.

Die Validität gibt an, ob der Test misst, was er messen soll (Grotjahn/Kleppin 2015: 50), und wird vorwiegend qualitativ bestimmt. Validität hat verschiedene Facetten – oft werden Inhaltsvalidität, Konstruktvalidität und kriterienbezogene Validität untersucht (Lienert/Raatz 1998: 10–11). Ein weiterer wichtiger Aspekt in der klassischen Testforschung ist die Augenschein-Validität (*face validity*) (Cronbach 1970: 182–183). Verschiedene Aspekte von Validität sollen in Bezug auf die Fachsprachen im nächsten Kapitel erörtert werden.

Die Inhaltsvalidität gibt an, inwiefern die Testaufgaben inhaltlich valide sind. Die Konstruktvalidität versucht zu beschreiben, welches Testkonstrukt der Test

abbildet und inwiefern dieses valide ist. Welche Konstrukte valide sind, ist in der Sprachtestforschung hoch umstritten (einführend hierzu Bachman/Palmer 1996: 21). Die kriterienbezogene Validität lässt sich als Korrelation mit einem Außenkriterium messen. In den Fachsprachenprüfungen kamen C-Tests als Außenkriterium zum Einsatz; daher kann auch diese Facette von Validität bestimmt werden. Die Augenschein-Validität schließlich gibt an, wie glaubhaft ein Testverfahren für die Testees ist. Sie ist kein unwichtiger Aspekt, denn nur wenn Prüflinge eine Prüfung sinnvoll und geeignet finden, werden sie ihre Bestleistungen abrufen.

Weitere Gütekriterien, die oft genannt werden, sind die Authentizität und Interaktivität (Bachman/Palmer 1996), aber auch Testökonomie, Testfairness und Normierung gehören zu den Nebengütekriterien (Moosbrugger/Kelava 2007: 18–24; Grotjahn/Kleppin 2015: 52–55). Im Rahmen dieses Aufsatzes wird es leider nur möglich sein, auf die Hauptgütekriterien einzugehen; die Nebengütekriterien müssen gesondert untersucht werden.

Bevor die Hauptgütekriterien in Bezug auf die Fachsprachenprüfungen näher untersucht werden, sollen noch relevante Forschungsergebnisse für die Untersuchung von mündlichen Sprachprüfungen grob referiert werden, da die Fachsprachenprüfungen per definitionem zu zwei Dritteln mündliche Prüfungen sind.

3.2 Mündliche Sprachprüfungen in der Forschung

Einen einführenden Überblick zu mündlichen Sprachprüfungen geben Sandlund, Sundqvist und Nyroos (2016) sowie Luoma (2004).

Sobald Sprachfähigkeit nicht mehr nur als Systemwissen, sondern vielmehr als kommunikative Kompetenz bzw. als Handlungsvermögen verstanden wird, müssen auch Prüfungsverfahren sowie Forschungsansätze grundlegend überarbeitet werden. Die Auswirkungen dieser kommunikativ-pragmatischen Wende, die seit 1970 auch in der deutschsprachigen Fremdsprachendidaktik und -forschung stattfindet, sind kaum zu unterschätzen (Piepho 1974; Bachman 1990; für die Analyse von Prüfungen sei besonders auf McNamara/Roevers 2006 verwiesen). Seit dieser fundamentalen Umwälzung haben völlig neue Prüfungsformate wie Gespräche mit einem Prüfer in der L2 (Oral Proficiency Interview, OPI) oder auch Gespräche von Prüflingen untereinander (*paired tests*) (Sandlund/Sundqvist/Nyroos 2016: 17–18) sowie strukturierte, halb strukturierte und offene Sprechaufgaben (Luoma 2004: 48–50) Eingang in mündliche Prüfungen gefunden. Sie werden nicht nur im schulischen oder universitären Alltag, sondern auch in diversen High-Stakes-Prüfungen regelmäßig eingesetzt. Nur beispielhaft seien hier das Gespräch mit einem Prüfer als Bestandteil der IELTS-Prüfungen (IELTS 2022: <https://www.ielts.org>) und die Gespräche von Prüflingen untereinander in

allen Prüfungen zu Goethe-Zertifikaten (Goethe-Institut online: <https://www.goethe.de/de/spr/kup/prf/prf.html>) erwähnt.

Nicht nur die Aufgabenformate, sondern auch die Bewertungskriterien und -verfahren in mündlichen Sprachprüfungen haben sich durch die Hinwendung zur Interaktion und Kommunikation als Prüfungsgegenstand stark verändert. So weist Luoma in ihrem Grundlagenwerk zur Bewertung des Sprechens (2004) darauf hin, dass die Spezifik der Mündlichkeit, wie die Bedeutung von Aussprache und Flüssigkeit, aber auch die spezifische Grammatik des Mündlichen und die pragmatische Ebene Eingang in die Konstruktion von mündlichen Prüfungen und Bewertungsskalen finden müssen. Dies wurde in zahlreichen mündlichen Prüfungen realisiert. Beispielsweise haben Spezifika mündlicher Interaktion wie *Flüssigkeit*, *Aussprache* oder auch *auf Argumente reagieren* etc. Eingang in das Bewertungsraster der mündlichen Teile bei den Prüfungen des Goethe-Instituts gefunden (Goethe-Institut 2018: 45).

Werden Gespräche als Prüfungsform eingesetzt, so müssen darüber hinaus zentrale Erkenntnisse der Gesprächsforschung berücksichtigt werden. Schon Wilhelm von Humboldt formulierte: „Es liegt aber in dem ursprünglichen Wesen der Sprache ein unabänderlicher Dualismus, und die Möglichkeit des Sprechens selbst wird durch Anrede und Erwiderung bedingt“ (Humboldt 1828: 23). Diese unauflösliche Dialogizität ist im 20. Jahrhundert insbesondere in der Conversation Analysis zentraler Forschungsgegenstand (vgl. auch Burkhardt 2018) geworden. Zahlreiche Analysen haben die Humboldt'sche These bestätigt, wonach Gespräche mehr als aus den einzelnen Beiträgen der Gesprächspartner zusammengesetzte Puzzle sind, sondern vielmehr Gemeinschaftswerke. Jedem noch so kleinen Gesprächselement ist neben der Präferenz des aktuell Sprechenden auch immer eine Spiegelung und Vorwegnahme der Reaktionen des Gesprächspartners inhärent, sodass im Grunde alle Gesprächsschritte gemeinsame Projekte sind (Stivers/Sidnell 2013: 2). Diese grundlegende Eigenschaft von Gesprächen muss bei jeder Prüfung, die Gespräche beinhaltet, berücksichtigt werden. Gespräche werden als Prüfungsaufgaben eingesetzt, um die Interaktionskompetenz und die Sprachfähigkeit in einem pragmatischen und handlungsorientierten Bezugsrahmen zu prüfen. Interaktionskompetenz ist jedoch in einem Gespräch keine Einzelleistung, sondern muss nach Galaczi so verstanden werden: „As a set of resources that reside not within an individual but are accomplished mutually and reciprocally by the participants in a discourse“ (Galaczi 2014: 553).

Ob ein Gesprächsziel erreicht wird, ist also immer von beiden Gesprächsteilnehmenden abhängig; somit kann der Gesprächserfolg keinesfalls alleiniges Bewertungskriterium sein. Der Umstand, dass Interviewende, die Gesprächspartner in mündlichen Prüfungen simulieren, in der Schwierigkeit der Interaktion differieren, konnte in größer angelegten Studien empirisch nachgewiesen werden

(Brown 2005). Besonders aufschlussreich ist auch die Studie von McNamara und Lumley (1997), in der das englischsprachige Pendant zu den Fachsprachenprüfungen untersucht wird, der Occupational English Test (OET). Die Autoren zeigen, dass sich die Kompetenz des bzw. der Prüfenden als Gesprächspartner auf die Bewertungen auswirkt (McNamara/Lumley 1997: 150).

Die Frage, welche beobachtbaren Bewertungskriterien im Gesprächsverhalten Prüflinge zuverlässig voneinander unterscheiden, wurde in der Forschungsliteratur zur Interaktionskompetenz (Interactional Competence, IC) unter anderem wie folgt beantwortet: Je fortgeschrittener Lernende sind,

- desto ausführlicher können Themen entfaltet werden und desto besser können Sprecher aufeinander Bezug nehmen (Gan 2010; Galaczi 2014),
- desto eher erfolgen Turn-Takings an *transition relevant places*, das heißt mit minimaler Überlappung (Galaczi 2014: 561),
- desto mehr Hörerrückmeldungen gibt es (ebd.) und
- desto besser können Lernende das eigene Anliegen rahmen, das heißt angemessen einleiten (Roever/Kasper 2018).

Zusammenfassend kann festgehalten werden, dass Prüfungsgespräche aufgrund ihrer dialogischen Natur besondere und hohe Anforderungen an alle Testentwickler und Testinstitutionen stellen. Da Gesprächsverlauf und Gesprächsergebnis in Prüfungsgesprächen, die wie in den Fachsprachenprüfungen zwischen Prüfling und Prüfer stattfinden, immer auch von Letzteren abhängen, ist es unumgänglich, spezifische Bewertungsskalen zu entwickeln, die die vielen Facetten der Mündlichkeit adäquat abbilden.

4 Hauptgütekriterien in den Fachsprachenprüfungen für Heilberufe

4.1 Objektivität

Wie oben schon dargelegt, können Prüfungen in Gesprächsform wohl nur schwerlich Durchführungsobjektivität für sich in Anspruch nehmen – zu zahlreich sind die Variablen, die jede Prüfungsdurchführung beeinflussen. Dennoch werden Prüfungsgespräche in verschiedenen High-Stakes-Prüfungen geführt. Die Durchführungsobjektivität wird dabei durch die folgenden Maßnahmen erhöht:

- standardisierte, mehr oder weniger geskriptete Begrüßungen, Einführungen und Aufgabenstellungen,

- standardisiertes Prüferverhalten, meist in Leitfäden festgehalten und in Prüferschulungen vermittelt,
- striktes Zeit- und Raummanagement, das allen Prüflingen gleiche Bedingungen vorgibt.

Diese Maßnahmen werden auch in den Fachsprachenprüfungen für Apotheker sowie Ärzte umgesetzt. In beiden Prüfungen existieren umfangreiche Leitfäden, Handreichungen und regelmäßige Prüferschulungen. Da die Fachsprachenprüfungen für Pflegekräfte noch nicht in eine reguläre Prüfungspraxis überführt wurden, steht diese Entwicklung hier noch aus.

Dennoch lässt sich in den transkribierten Prüfungsgesprächen deutlich erkennen, dass Prüfende sehr unterschiedlich mit Problemen im Gespräch umgehen. Hier soll nur kurz auf ein zentrales Problem eingegangen werden: auf den Umgang mit unverständlichen Äußerungen. Die Herstellung von Verständnis ist ein zentrales Anliegen in allen Gesprächen; daher spielen Reparaturen in der Conversation Analysis eine große Rolle (Hutchby/Wooffitt 2008: 57). Einerseits bedrohen besonders fremdinitiierte Fremdreparaturen die Gesichtswahrung und werden daher oft vermieden (Brown/Levinson 1987: 65); andererseits ist die Verständnissicherung gerade im Gespräch mit Patienten essenziell. In den beobachteten und transkribierten Prüfungsgesprächen differiert der Umgang mit unverständlichen Passagen erheblich; hier seien nur zwei unterschiedliche Reaktionen aufgezeigt. Das erste Beispiel stammt aus einem Apotheker-Patienten-Gespräch; es geht um die Abgabe von zwei verschiedenen Lungenmedikamenten. Das erste Medikament wurde schon erklärt, es folgt diese Nachfrage:

Tabelle 3: Transkriptausschnitt, FSP für Apothekerinnen und Apotheker, Apotheker-Patienten-Gespräch, Kandidat 15, 2018

38	P	[und] wann und wann soll ich dann das andere nehmen?
39	Ap	<i>bei wenn sie atemnot haben. also bei bedarf wenn sie verkümmte lungen haben.</i>
40	P	aha
41	Ap	<i>das nur entzündliche probleme.(1) bei einzündlichen. (2) #00:03:12-6#</i>
42	P	aber ich hab doch nur asthma #00:03:13-8#

Die Auskünfte in den Segmenten 39 und 41 sind unverständlich, es erfolgt aber keine Korrektur und auch keine Nachfrage. Der weitere Gesprächsverlauf zeigt, dass insgesamt weitere zehn Minuten damit vergehen, dass der Apotheker die Anwendung des Medikaments zu erklären versucht.

Die entgegengesetzte Strategie einer anderen Prüferin ist, bei jeder unverständlichen Äußerung um Klärung zu bitten, mit anderen Worten eine Selbstreparatur zu initiieren. Aufgrund der Nachfrage in Segment 167 („das ist für mich kinesisch“) bekommt der Prüfling die Chance, seine Aussage zu reparieren und zu erklären:

Tabelle 4: Transkriptausschnitt, FSP für Ärztinnen und Ärzte, Arzt-Patienten-Gespräch, Kandidat 1, 2019

166	A	<i>ä also sprunggelenkdediskropation [glaub ich] #00:22:21-6#</i>
167	P	<i>[moment moment] moment was heisst das ich das ist für mich kinesisch. #00:22:25-8#</i>
168	A	<i>also der sprunggelenk ist ein gelenk der ein bisschen kompliziert ist das sind sehr viele knochen drin. und ä wenn sie eine verletzung haben also wie sie jetzt ä mit mit dem stein gestolpert sind, dann kann ein knochen also rausragen. oder er ragt rich nicht raus aber die gelenkte die diesen gelenk formen können abreißen. #00:22:53-2#</i>
169	P	<i>hm_hm #00:22:54-4#</i>
170	A	<i>und das kann auch das verursachen. im schlimmsten fall kann man auch eine fraktur sein, aber sie sagen sie haben kein blut gesehen, also dass es rausgeht, oder? #00:23:05-3#</i>
171	P	<i>nein, das nicht #00:23:06-5#</i>

Dies sind zwei kleine Ausschnitte, die verdeutlichen, dass das Prüfungsformat der mündlichen Prüfung im Gespräch große Schwierigkeiten bei der Erreichung von Durchführungsobjektivität impliziert. Hier sind Prüferschulungen, wie sie die Apotheker- und Ärztekammern regelmäßig durchführen, von hohem Wert. In den Fachsprachenprüfungen für Apotheker wurde neben dem Prüfungsgespräch auch ein kurzer Monolog zu einem Thema aus dem Berufsalltag eingeführt – hier kann die mündliche Kompetenz objektiver als im Gespräch geprüft werden.

Ein weiterer Faktor, der die Durchführungsobjektivität einschränkt und eine genauere Analyse verdiente, als sie hier geleistet werden kann, ist der Umstand, dass in den Fachsprachenprüfungen für Ärzte sowie für Apotheker inzwischen zahlreiche Szenarien zum Einsatz kommen. Somit soll verhindert werden, dass die Kandidaten Prüfungsgespräche im Detail vorher auswendig lernen können, denn in verschiedenen Internetforen, Blogs und Facebook-Gruppen werden mittlerweile in großem Umfang und mit erstaunlicher Detailtreue Szenarien erinnert und aufbereitet.

Es müsste nunmehr erforscht werden, ob diese verschiedenen Szenarien unterschiedlich schwierig sind. Hierfür bräuchte man eine große Menge an Prüfungsprotokollen, um dann die Verfahren der probabilistischen Testforschung einsetzen zu können, ähnlich wie bei Brown (2005). Im schriftlichen Teil ist die

Durchführungsobjektivität insofern gegeben, als alle Prüflinge unter denselben Konditionen ein Dokument erstellen müssen. Aber auch hier bleibt unklar, inwiefern Szenarien unterschiedlich schwierig sind, da hierfür systematisch Itemschwierigkeiten bestimmt und miteinander verglichen werden müssten, was im Zuge der Entstehung dieser neuen Prüfungen (noch) nicht der Fall ist.

Zur Objektivität einer Prüfung gehört auch immer die Objektivität der Auswertung. Wie schon oben in Abschnitt 3.1 dargestellt, ist die Auswertungsobjektivität bei geschlossenen Aufgabenformaten mehr oder weniger gegeben, während sie bei so offenen Formaten wie den hier diskutierten Gesprächen nur durch konsequente Bewerter Schulungen angestrebt werden kann. Zu allen hier diskutierten Prüfungen gehören detaillierte Bewertungsraster, die auch Skalen zur grammatischen Korrektheit, zur Flüssigkeit, zum Wortschatzumfang, aber auch zur Gesprächsführungskompetenz und zum Verstehen enthalten. Viele Skalen orientieren sich an den Deskriptoren des GER, was der Konstruktvalidität der Prüfung dient. Die Raster für Ärzte sowie Apotheker sind klar strukturiert und in Schulungen gut vermittelt, das Raster für die Pflegekräfte wurde noch nicht in der Praxis im großen Umfang eingesetzt, hat sich aber in der Pilotierung bewährt. Die Raster sind ein wichtiger Bestandteil der Prüfung und sollten weiter verbessert werden. Dabei sollten weitere Aspekte der Sprachkompetenz wie Aussprache, Kohäsion und Kohärenz Eingang finden. Auch die oben referierten Forschungserkenntnisse zu Hörerrückmeldungen, Rahmungen sowie Überlappungen als Indikatoren für Sprachkompetenz in mündlichen Prüfungsgesprächen könnten nach sorgfältiger Pilotierung Eingang in neue Bewertungsskalen finden.

Insgesamt lässt sich festhalten, dass es im vorgegebenen Testformat schwer ist, Durchführungs- und Auswertungsobjektivität zu erreichen. Wie so oft im Kontext von Sprachprüfungen, wird auch hier der Validität Vorzug vor der Objektivität gegeben. Prüferschulungen zu allen Aspekten der Prüfung spielen eine große Rolle und weitere Forschungsarbeit ist erforderlich, um die unterschiedlichen Eigenschaften von Prüfenden und die Szenarien besser abschätzen und kalibrieren zu können.

4.2 Reliabilität

Wie oben dargelegt, gibt die Reliabilität Auskunft über die Zuverlässigkeit der Messung. Sie konnte für die Fachsprachenprüfungen für Ärzte ($n = 33$) sowie Pflegekräfte ($n = 27$) anhand der internen Konsistenz mit Cronbachs α bestimmt werden. Diese Stichprobe ist allerdings so klein, dass sie natürlich nicht repräsentativ ist und die nun folgenden Berechnungen nur mit Vorbehalt betrachtet werden können. Dennoch ist dies die einzige Methode, um die Reliabilität überhaupt zu

bestimmen. In der Fachsprachenprüfung für Ärzte kam ein Bewertungsbogen mit 13 Skalen, die jeweils vierschrittig waren, zum Einsatz; für die Fachsprachenprüfung für Pflegekräfte wurde ein Bewertungsbogen mit 18 ebenfalls vierschrittigen Skalen entwickelt. Alle Skalen wurden in eine Matrix überführt und anhand dieser Werte die Korrelation der Skalen untereinander berechnet. Ist Cronbachs α höher als 0,9, wird die Reliabilität als sehr gut eingeschätzt (Moosbrugger/Kelava 2007: 129).

Für die hier analysierten Fachsprachenprüfungen für Ärztinnen und Ärzte konnte Cronbachs α mit 0,95 bestimmt werden, was ein außerordentlich hoher Wert ist. Weiterhin erwiesen sich alle Skalen als trennscharf. Cronbachs α konnte auch durch Auslassung einer Skala nicht gesteigert werden. Für die Fachsprachenprüfungen für Pflegekräfte wurde Cronbachs α mit 0,97 errechnet. Auch hier waren alle Skalen trennscharf und Cronbachs α konnte nicht gesteigert werden. Zumindest für diese kleine Stichprobe kann bei allen Einschränkungen also von einer sehr hohen Zuverlässigkeit gesprochen werden.

4.3 Validität

Dass die hier vorgestellten Prüfungsaufgaben, wie das Anamnesegespräch mit einem Patienten, die Erklärung einer Medikamentenabgabe, die Anfertigung eines Pflegeprotokolls oder auch die Diskussion von Verdachts- und Differentialdiagnosen mit einem Kollegen, inhaltlich valide und authentisch sind, steht außer Zweifel. Alle Aufgaben in den Fachsprachenprüfungen entstammen dem Berufsalltag von Ärzten, Apothekern respektive Pflegekräften. Es ist ein großer Vorzug aller hier genannten Prüfungen, dass sie stets von Berufspraktikern erstellt wurden. Aufgrund der hohen Zahl an Prüfungen wurden von den Kammern mittlerweile sehr viele Prüfungsszenarien entwickelt.

Gleichzeitig haben sich vor allem im Internet verschiedenste Formen des Austauschs zu den genauen Prüfungsinhalten etabliert, ebenso wie zahlreiche Vorbereitungskurse privater Sprachinstitute auf die Fachsprachenprüfungen. Infolgedessen konnte oft beobachtet werden, dass Kandidatinnen und Kandidaten ganze Gesprächspassagen auswendig gelernt hatten und wie im folgenden Beispiel die Strukturen zum Teil logisch falsch verwendeten. In dieser Fachsprachenprüfung kommt die Patientin nach einem Unfall zum Arzt, der das Anamnesegespräch führt. Die Patientin hat den Unfallhergang bereits ausführlich geschildert, als der Arzt fragt.

Tabelle 5: Transkriptausschnitt, Fachsprachenprüfung für Ärztinnen und Ärzte, Kandidat 2, 2019

98	A	<i>ok ä alles klar. äm ä wo be also diese beschwerden sind bei ihnen erstes mal oder wurde das schon mal passiert diese diese dieselben beschwerden haben sie schon mal gehabt oder ist das erstes mal #00:07:41-5#</i>
99	P	moment jetzt jetzt hab ich das jetzt nicht verstanden, welche beschwerden
100	A	<i>ä ich meine also diese schwellungen, schmerzen ä #00:07:49-6#</i>
101	P	<<f<aber ich bin doch UMgeknickt!
102	A	<i>[ok]</i>

Dieser Austausch ist wohl sehr prüfungsspezifisch und hat mit dem beruflichen Alltag vermutlich eher weniger gemein. Eine weitere Einschränkung der Inhaltsvalidität ergibt sich zumindest in den Fachsprachenprüfungen für Ärzte sowie Pflegekräfte aus dem Umstand, dass keine körperliche Untersuchung vorgenommen wird und die berichteten körperlichen Zustände nur fiktiv sind. Das Gespräch zwischen Apotheker und Patient ist davon eher nicht betroffen, da hier das Medikament im Zentrum steht. In den Fachsprachenprüfungen für Ärzte wird aber zum Beispiel trotz manchmal behaupteter starker Schmerzen das Anamnesegespräch ruhig fortgesetzt, sodass sich auch hier die Inhaltsvalidität anzweifeln lässt.

Schließlich sollte in einer tiefer gehenden Analyse der Umstand beachtet werden, dass in dem spezifischen Setting der Fachsprachenprüfungen für Ärzte sowie für Apotheker diverse Machtverhältnisse einander entgegenstehen: Der Arzt respektive Apotheker hat eine größere diskursive Macht als der Patient; der (Simulations-)Patient ist aber andererseits L1-Sprecher des Deutschen und Mitglied der Prüfungskommission, während Arzt und Apotheker L2-Sprecher und Prüflinge sind. Diese schwierige Situation gefährdet die Inhaltsvalidität und ließe sich zum Beispiel dadurch zum Teil beheben, dass der Simulationspatient respektive der Simulationskollege nicht der Prüfungskommission angehört und nicht mitbewertet. Professionelle Schauspielerinnen und Schauspieler, wie sie auch im Medizin- oder Pharmaziestudium bereits eingesetzt werden, wären der Objektivität der Prüfung dienlich.

Dennoch kann im Großen und Ganzen die Inhaltsvalidität bestätigt werden. Verweisen jedoch diese Inhalte auf ein valides Konstrukt von Sprachkompetenz von Heilberufen? Hier fällt auf, dass wichtige Dimensionen des Konstrukts, wie zum Beispiel Lesekompetenz, aber auch grammatische Kompetenz und phonetische Kompetenz, nicht oder nur am Rande berücksichtigt werden. Die oben genannten GMK-Beschlüsse erfassen nur einen Teil der komplexen sprachlichen Wirklichkeit in Heilberufen; die Konstruktvalidität müsste also durch weitere Prüfungsaufgaben und -dimensionen gestärkt werden.

Schließlich konnten in den Projekten mit der Ärztekammer und dem Nord-deutschen Pflegezentrum (NDZ) allgemeinsprachliche C-Tests eingesetzt werden. Ähnlich wie bei der Berechnung der Reliabilität muss die geringe Stichprobengröße berücksichtigt werden; dennoch lassen sich so allererste Aussagen zu einer kriterienbezogenen Validität treffen, da der C-Test ein Außenkriterium darstellt. Die genauen Prüfungsergebnisse waren von 33 Ärztinnen und Ärzten und 27 Pflegekräften bekannt, sodass zunächst die Normalverteilung der Prüfungsergebnisse und der C-Test-Ergebnisse berechnet werden konnte. Da alle drei Reihen Normalverteilung nach Shapiro-Wilk aufwiesen, konnte jeweils die Korrelation nach Pearson zwischen dem C-Test und dem FSP-Ergebnis bestimmt werden. Dies führte zu folgenden Ergebnissen:

Tabelle 6: Korrelationen nach Pearson zwischen C-Test und Fachsprachenprüfungen

	n	Korrelation nach Pearson, r
Fachsprachenprüfung für Ärzte (März 2021)	33	0,5512
Fachsprachenprüfung für Pflegekräfte (September 2021)	27	0,3280

Beide Fachsprachenprüfungen korrelieren also positiv mit einem Außenkriterium, das in beiden Projekten dasselbe war: Die Ergebnisse der Ärztinnen und Ärzte korrelieren in einem starken Maß mit den C-Test-Ergebnissen, die Ergebnisse der Pflegekräfte immer noch moderat. Die Pflegeprüfung ist noch nicht etabliert. Auch die Prüfenden und die Bewertenden haben diese Prüfung zum ersten Mal abgenommen, während die Fachsprachenprüfung für Ärzte bereits ein etabliertes Format ist. Die Differenz lässt sich auf diesen Umstand zurückführen, denn in den Ärzte- respektive Apothekerkammern finden bereits regelmäßige Bewerberschulungen statt – anders als im Pilotprojekt für die Pflegekräfte.

Abschließend kann konstatiert werden, dass Inhalts- und Konstruktvalidität weiter erforscht werden müssen. Die kriterienbezogene Validität scheint aber zumindest bei den Fachsprachenprüfungen für Ärzte und Pflegekräfte gegeben. Auch die Augenschein-Validität ist hoch, wie Beobachtungen und Gespräche mit Prüflingen ergaben. Hier gibt es keinerlei Zweifel an den Prüfungsaufgaben oder der Prüfungsdurchführung, vor allem weil die Aufgaben inhaltsvalide erscheinen. Dies führt zu einer sehr hohen Motivation der Prüflinge.

5 Zusammenfassung und Ausblick

Die Fachsprachenprüfungen für Heilberufe stellen relativ neue Prüfungen in Deutsch als Fremdsprache dar. Dass diese Prüfungen Testgütekriterien erfüllen, ist nicht nur aus testtheoretischer Sicht erstrebenswert, sondern in höchstem Maße relevant für die Berufsausübung und letzten Endes für die Patientensicherheit und das Patientenwohl in Apotheken, Krankenhäusern, Ambulanzen etc.

Hier ergibt sich ein reiches und wichtiges Forschungsfeld für die angewandte DaF-Forschung. In drei kleineren Projekten konnten erste Einblicke in die Fachsprachenprüfung für Apotheker sowie für Ärzte und die gerade neu entwickelte Fachsprachenprüfung für Pflegekräfte gewonnen werden. Da die Prüfungen zeitaufwendige Einzelprüfungen sind und der Zugang zum Feld schwierig ist und bleibt, kommen zur Erforschung vor allem qualitative Methoden, ergänzt um einfache quantitative Methoden der klassischen Testtheorie, zum Einsatz. Wünschenswert wäre auf jeden Fall die Erhebung und statistische Auswertung einer größeren Menge an Prüfungen, um so auch qualifiziertere Aussagen zur Durchführungsobjektivität und zur Reliabilität treffen zu können. In diesem Aufsatz konnte nur vermutet werden, dass die Durchführungsobjektivität unter den verschiedenen Prüfenden und den verschiedenen Szenarien leidet, wie in Abschnitt 4.1 auch anhand von Beispielen gezeigt werden konnte. Wie bei allen mündlichen Prüfungen, die in Gesprächsform stattfinden, sind die verschiedenen Aspekte der Objektivität nur schwer zu erreichen. Die Reliabilität hingegen ist dem ersten Anschein nach durchgehend hoch, ebenso wie verschiedene Aspekte der Validität. Da die Fachsprachenprüfungen vor allem mündliche Prüfungen sind, konnten erfolgreich Verfahren der Conversation Analysis angewandt werden. In diesem Rahmen konnte das Verfahren nur angedeutet werden, aber letztendlich sind Prüfungstranskripte bei der Diskussion von Validität und Objektivität von großem Nutzen. Darüber hinaus ermöglichen sie, die diskursive Praxis, die in diesen Prüfungsgesprächen sowohl eingeübt als auch ausgeübt wird, zu erkennen. Im Transkript lassen sich Diskursrollen und Machtverteilungen erkennen. Transkripte ermöglichen zudem, Prüfungs- und Bewertungspraktiken im Detail nachzuvollziehen und damit auch das eigene Fach kritisch zu reflektieren. Aus Platzgründen ist es in diesem Kontext nicht möglich gewesen, diesen Aspekt zu vertiefen, aber in Folgevorhaben wäre es durchaus lohnend und wünschenswert, den sozialen Kontext dieser Prüfungen anhand der Transkripte näher zu beleuchten.

Prüfungen sind ein „Schibboleth“ (McNamara 2012) – sie fungieren als abschließende Praxis, in diesem Fall regeln sie den Zutritt zum Arbeitsmarkt für Angehörige verschiedener Heilberufe. Ein so komplexes Konstrukt wie Sprachkompetenz ist immer auch sozial definiert und erschafft seinerseits auch wieder soziale Praktiken. Welches Schibboleth Apotheker und Ärzte oder Pflegekräfte

aus dem Ausland genau und in welcher Weise aussprechen müssen, ist eine hoch relevante Frage und Perspektive, die späterer Forschung vorbehalten bleibt.

Literaturverzeichnis

- ABDA (2021a): *Als ausländischer Apotheker in Deutschland arbeiten*. Online: <https://www.abda.de/fuer-apotheker/berufsausuebung/anererkennung-der-apothekerdiplo-me/als-auslaendischer-apotheker-in-deutschland-arbeiten/> (16.11.2021).
- ABDA (2021b): *Die Apotheke: Zahlen, Daten, Fakten*. Internetbroschüre. Online: https://www.abda.de/fileadmin/user_upload/assets/ZDF/ZDF21/ABDA_ZDF_2021_Broschuere.pdf (17.11.2021).
- Bachman, Lyle F. (1990): *Fundamental considerations in language testing*. Oxford: University Press.
- Bachman, Lyle F.; Palmer, Adrian S. (1996): *Language testing in practice: Designing and developing useful language tests*. Oxford: University Press.
- Brown, Annie (2005): *Interviewer variability in Oral Proficiency Interviews*. Frankfurt am Main: Peter Lang.
- Brown, Penelope; Levinson, Stephen (1987): *Politeness: Some Universals in Language Use*. Cambridge: University Press.
- Bundesagentur für Arbeit (2021): *Berichte Blickpunkt Arbeitsmarkt Fachkräftengpassanalyse 2020*. Nürnberg. Online: https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html;jsessionid=E0C97FFD38A433D4AFA2F410ACA40CD5?nn=20626&topic_f=fachkraefte-engpassanalyse (13.11.2021).
- Bundesärztekammer (2020): *Ärzttestatistik 2020*. Online: <https://www.bundesaerztekammer.de/ueber-uns/aerzttestatistik/auslaendische-aerztinnen-und-aerzte/> (10.06.2021).
- Burkhardt, Armin (2018): „Der Dialog-Begriff bei Wilhelm von Humboldt und seine Folgen“. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 48 (4), 717–740.
- Cronbach, Lee J. (1970): *Essentials of psychological testing*. 2. Auflage. New York: Harper & Row.
- Fulcher, Glenn; Davidson, Fred (2016) *The Routledge Handbook of Language Testing*. London: Routledge.
- Galaczi, Evelina D. (2014): „Interactional Competence across Proficiency Levels: How do Learners Manage Interaction in Paired Speaking Tests?“. In: *Applied Linguistics* 35 (5), 553–574.
- Gan, Zhengdong (2010): „Interaction in Group Oral Assessment: A case study of higher- and lower-scoring students“. In: *Language Testing* 27 (4), 585–602.
- Gesundheitsministerkonferenz (2014): *Eckpunkte zur Überprüfung der für die Berufsausübung erforderlichen Deutschkenntnisse in den akademischen Heilberufen*. Online: https://www.gmkonline.de/documents/TOP73BerichtP_Oeffentl_Bereich.pdf (15.10.2020).
- Gesundheitsministerkonferenz (2019): *Eckpunktepapier zur Überprüfung der für die Berufsausübung erforderlichen Deutschkenntnisse in den Gesundheitsfachberufen*. Online: https://www.gmkonline.de/documents/anlage-top86_92gmk-eckpunkte_1570622947.pdf (10.11.2021).
- Goethe-Institut e.V. (Hrsg.) (2018): *Materialien zur Prüfung Goethe-Zertifikat B2*. Online: https://www.goethe.de/pro/relaunch/prf/materialien/B2/b2_modellsatz_erwachsene.pdf (14.07.2022).
- Grotjahn, Rüdiger; Kleppin, Karin (Hrsg.) (2015): *Prüfen, Testen, Evaluieren*. Berlin: Klett-Langenscheidt.

- Humboldt, Wilhelm von (1828): Ueber den Dualis: gelesen in der Akademie der Wissenschaften am 26. April 1827. Berlin: Druckerei der Königlichen Akademie der Wissenschaften.
- Hutchby, Ian; Wooffitt, Robin (2008): *Conversation analysis: Principles, practices and applications*. 2. Auflage. Cambridge: Polity Press.
- IELTS – The International English Language Testing System (2022): Online: <https://www.ielts.org/> (16.08.2022).
- Lienert, Gustav A.; Raatz, Ulrich (1998): *Testaufbau und Testanalyse*. 6. Auflage. Weinheim: Beltz.
- Luoma, Sari (2004): *Assessing Speaking*. Cambridge: University Press.
- Marburger Bund (2020): *Deutschkenntnisse – Anforderungen in den Bundesländern für die Approbationserteilung: German language requirements*. Online: <https://www.marburger-bund.de/bundesverband/service/auslaendische-aerzte/foreign-physicians/anforderungen-deutschkenntnisse> (16.11.2021).
- McNamara, Tim (2012): „Language Assessments as Schibboleths: A Poststructuralist Perspective“. In: *Applied Linguistics* 33 (5), 564–581.
- McNamara, Tim; Lumley, Tom (1997): „The effect of interlocutor and assessment mode variables in overseas assessment of speaking skills in occupational settings“. In: *Language Testing* 14 (2), 140–156.
- McNamara, Tim; Roever, Carsten (2006): *Language Testing: The Social Dimension*. Oxford: Blackwell.
- Moosbrugger, Helfried; Kelava, Augustin (Hrsg.) (2007): *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Nier, Hedda (2017): *Fachkräftemangel in Gesundheits- und Pflegeberufen*. Online: <https://de.statista.com/infografik/11447/fachkraeftemangel-in-gesundheits-und-pflegeberufen/> (13.11.2021).
- Piepho, Hans-Eberhard (1974): *Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht*. Limburg: Frankonius.
- Piepho, Hans-Eberhard (2003): *Lerneraktivierung im Fremdsprachenunterricht: „Szenarien“ in Theorie und Praxis*. Hannover: Schroedel.
- Radtke, Rainer (2021): *Ausländeranteil an sozialversicherungspflichtig Beschäftigten in der Pflege in Deutschland nach Pflegeart im Jahr 2020*. Online: <https://de.statista.com/statistik/daten/studie/1029896/umfrage/auslaenderanteil-an-pflegekraeften-in-deutschland-nach-pflegeart/> (12.11.2021).
- Roever, Carsten; Kasper, Gabriele (2018): „Speaking in turns and sequences: Interactional competence as a target construct in testing speaking“. In: *Language Testing* 35 (3), 331–355.
- Sandlund, Erica; Sundqvist, Pia; Nyroos, Lina (2016): „Testing L2 Talk: A Review of Empirical Studies on Second-Language Oral Proficiency Testing“. In: *Language and Linguistics Compass* 10 (1), 14–29.
- Schön, Almut; Johnson, Natalia; Zimmermann, Kerstin (2014): „Schwellen erkennen – verschiedene Möglichkeiten zur Festlegung von *cut scores* in C-Tests“. In: Mügge, Regina (Hrsg.): *Gekonnt, verkannt, anerkannt – Sprachen im Bologna-Prozess*. Dokumentation der 27. Arbeitstagung des AKS 2012. Bochum: Verein z. Förd. d. AKS, 71–88.
- Stivers, Tanya; Sidnell, Jack (2013): „Introduction“. In: Stivers, Tanya; Sidnell, Jack (Hrsg.): *The handbook of conversation analysis*. Malden Mass.: Wiley-Blackwell, 1–8.
- Trim, John; North, Brian; Coste, Daniel; Sheils, Joseph (2001): *Gemeinsamer Europäischer Referenzrahmen für Sprachen*. Berlin: Langenscheidt.

Biographische Angaben

Almut Schön

ist Geschäftsführerin der Zentraleinrichtung Moderne Sprachen (ZEMS) an der TU Berlin. Sie studierte in Berlin und Uppsala Germanistik und Skandinavistik und promovierte an der TU Berlin zur Arzt-Patienten-Interaktion als L2-L1-Interaktion. Sie forscht insbesondere zur Bedarfsanalyse, zur Gesprächsanalyse, zu Sprachtests und zur Diskursanalyse.