

Testen und Prüfen im Bereich Deutsch als Fach- und Fremdsprache

Zusammenfassung: Dieser Beitrag beschäftigt sich mit Testen und Prüfen im Bereich Deutsch als Fremd- und Fachsprache. Zu Beginn erfolgt eine Einführung in das Themengebiet, indem zentrale Begriffe besprochen und testtheoretische Grundlagen erörtert werden. Exemplarisch werden dann qualitätssichernde Maßnahmen bei einem Sprachtestformat – der *Deutschen Sprachprüfung für den Hochschulzugang* (DSH) – beschrieben, wie sie am Standort Universität Duisburg-Essen (UDE) umgesetzt werden. Daran schließt sich eine strukturierte Auflistung bestehender Sprachprüfungen an, die einen Überblick über die Angebote in der Prüfungslandschaft gibt. Die Einteilung in allgemein-sprachliche und zweckorientierte Prüfungen soll dabei die Funktionen der Prüfungen hervorheben. Abschließend werden perspektivisch Möglichkeiten für diesen Bereich diskutiert.

Schlagwörter: Sprachprüfung, DSH, GER, Qualitätssicherung, allgemeinsprachliche Prüfungen, berufsbezogene Prüfungen, fachsprachliche Prüfungen

- 1 Einleitung
- 2 Testen und Prüfen im Bereich Sprache?
- 3 Prüfungsentwicklung am Beispiel der DSH am Standort Universität Duisburg-Essen
- 4 Prüfungen im Bereich Deutsch als Fach- und Fremdsprache
- 5 Ausblick
- 6 Literatur

1 Einleitung

Der Erhebung des Auswärtigen Amtes mit dem Titel „Deutsch als Fremdsprache weltweit“ zufolge lernten 15,45 Millionen Menschen im Jahr 2020 Deutsch als Fremdsprache. Deutsch gehört zu den am häufigsten gelernten Fremdsprachen der Welt. Die Gründe für das erhebliche Interesse an der deutschen Sprache sind vielfältig, z. B. der Wunsch, in einem deutschsprachigen Land zu studieren und zu forschen. Mit qualifizierten Deutschkenntnissen werden zudem bessere berufliche Chancen verbunden, denn diese werden sowohl für die Arbeit bei einem deutschen Unternehmen im Ausland als auch für die Arbeitssuche und eine spätere Einstellung in deutschsprachigen Ländern zumeist vorausgesetzt. Für den Zugang zum Arbeits- und Ausbildungsmarkt ist das Bestehen einer entsprechenden Sprachprüfung zumeist notwendig, weshalb das Testen und Prüfen im Bereich Sprachen von besonderer Bedeutung für die Betroffenen ist. Zunächst soll daher geklärt werden, was unter Testen und Prüfen im Bereich Sprachen verstanden wird und welche Aspekte in diesem Zusammenhang relevant sind.

2 Testen und Prüfen im Bereich Sprache

Das Testen und Prüfen im Bereich Deutsch als Fach- und Fremdsprache umfasst die Leistungsmessung mithilfe eines methodischen Verfahrens zur Feststellung der sprachlichen Fähigkeiten. Die Begriffe Testen und Prüfen werden zumeist synonym verwendet (so auch in diesem Beitrag), aber zuweilen auch nach dem Grad ihrer Formalität unterschieden. Oft wird dabei der Begriff Testen im informellen und der Begriff Prüfen im formellen Rahmen genutzt. Beiden Begriffen ist eigen, dass sprachliche Kompetenzen erfasst und beurteilt werden und dadurch die Leistungsfähigkeit oder der Kenntnisstand des Geprüften festgestellt wird (vgl. Perlmann-Balme 2001, 995).

Prüfungen werden unterschiedlichen Kategorien zugeordnet, je nachdem, welches Ziel mit der Prüfung erreicht werden soll bzw. welche Funktion sie hat. In Grotjahn (2005) findet sich beispielsweise eine Unterscheidung zwischen der Selektions-, der Diagnose- und der Zertifizierungsfunktion. Bei der Selektionsfunktion wird bei Erreichen der Anforderungen durch die Feststellung eines festgelegten Sprachstandes der Zugang zu einer bestimmten Institution gewährt. Bei der Diagnosefunktion wird der prozesshafte Spracherwerb fokussiert, um den Förderbedarf eines Lernenden zu ermitteln. Bei der Zertifizierungsfunktion werden Sprachkenntnisse zertifiziert, wodurch der Zugang zu einer Institution ermöglicht wird. Dies trifft auch auf die Selektionsfunktion zu, weshalb beide Funktionen wesentliche Überschneidungen in Bezug auf die Auswirkung aufweisen. Eine andere Klassifizierung nimmt Glaboniat (2010, 335) vor, die drei Hauptarten von Sprachtests unterscheidet: Diagnostische Tests, Lernfortschrittstests und Proficiency Tests. Diagnostische Tests dienen nach Glaboniat vor allem als Einstufungstests, die als Ziel nicht vorrangig die Sprachförderung haben, sondern die Festlegung sprachlicher Kompetenzen (vgl. Settineri/Jeuk 2019, 3 in Bezug auf die Bedeutungserweiterung des Terminus *Diagnose*). Lernfortschrittstests werden von Glaboniat als Achievement Test oder Leistungstest bezeichnet. In der Regel werden diese Prüfungen eingesetzt, um den entsprechenden Förderbedarf festzustellen, was der Diagnosefunktion nach Grotjahns Einteilung gleichkommt. Proficiency Tests werden als Feststellungs-, Niveau- oder Qualifikationsprüfungen konkretisiert. Diese Prüfungen haben einen formelleren Charakter als die beiden anderen Kategorien. Eine weitere Differenzierung der Proficiency Tests wird in Abschnitt 4 vorgenommen, in dem allgemeinsprachliche und zweckgebundene Prüfungen erläutert werden.

Je nach Funktion wird für die Messung der Leistung eine entsprechende Bezugsnorm herangezogen. Um den Lernfortschritt zu ermitteln oder eine Diagnose der Leistung vorzunehmen, wird in der Regel entweder der Fortschritt anhand der persönlichen Entwicklung (individuelle Norm) oder die Leistung im Vergleich zur Gesamtgruppe (kollektive Norm) bewertet. Durch den Vergleich wird eine relative Leistung festgestellt. Ein Beispiel für die Anwendung der kollektiven Norm sind Vokabeltests, bei denen die Notengebung des Einzelnen über die vergleichende Leistung zur Gruppe erfolgt (vgl. Arras 2007, 70). Hierbei wird von einer *normorientierten Evaluation* ge-

sprochen. Da diese Diagnose in Abhängigkeit von der jeweiligen Gruppenstärke unterschiedlich ausfallen kann, wird in diesen Fällen von nicht-standardisierten Tests gesprochen. Eine weitere Kategorie stellt die *kriterienbezogene Orientierung* dar, bei der zuvor festgelegte Kriterien als Maßstab dienen. Eine Kriterienorientierung dient als erster und wichtigster Ansatz, um eine Standardisierung der Prüfung zu erreichen. Die Voraussetzung dafür ist, dass ein einheitlicher Maßstab die Grundlage der Bewertung darstellt. Dieser ist insbesondere bei Tests mit Selektions- und Zertifizierungsfunktion bzw. bei Proficiency Tests notwendig, da die Ergebnisse dieser Prüfungen eine hohe Relevanz für die Prüfungsteilnehmenden haben und teils zukunftsbestimmende Konsequenzen nach sich ziehen können. Je bedeutender die Konsequenzen für den Prüfungsteilnehmenden sind, desto formeller sind die Prüfungen. Sehr formelle Prüfungen wie Proficiency Tests werden als High-Stakes-Prüfungen bezeichnet. Für sie ist eine Kriterienorientierung unerlässlich. Die für diese Prüfungen herangezogenen Kriterien beschreiben den sprachlichen Zustand, der auf einem bestimmten Niveau erwartet wird. Sowohl im Bereich Deutsch als Fremdsprache als auch im Bereich Deutsch als Fachsprache ist hierfür der gängige Bezug der *Gemeinsame Europäische Referenzrahmen für Sprachen: lernen, lehren und beurteilen* (GER).

Der GER ist in der deutschen Erstausgabe im Jahre 2001 erschienen. Er ist ein vom Europarat ins Leben gerufenes Projekt mit dem Ziel, eine gemeinsame Plattform im Umgang mit Sprachen zu schaffen. Im GER sind kommunikative Kompetenzen von Sprachlernenden als sogenannte Kann-Beschreibungen formuliert. Eine kommunikative Kompetenz zeigt sich durch die Befähigung, sprachliche Handlungen „kontextspezifisch und sozial angemessen zu produzieren und zu rezipieren“ (Studer 2010, 1265). Die Kann-Beschreibungen sind Positivformulierungen kommunikativer Kompetenzen. In diesen Beschreibungen sind die vier sprachlichen Fertigkeiten (Lesen, Hören, Schreiben, Sprechen) eingeteilt in sechs Kompetenzniveaustufen (A1, A2, B1, B2, C1, C2) in einem Skalierungssystem dargestellt. Die einzelnen Niveaustufen werden durch Deskriptoren, die die jeweiligen Sprachstandserwartungen beschreiben, veranschaulicht (vgl. Quetz/Schieß/Skörries/Goethe-Institut 2013, 8). Eine zentrale Veränderung, die der GER gefestigt hat, ist die Handlungsorientierung. Das handlungsorientierte Testen richtet die Aufmerksamkeit in der Testerstellung auf die kommunikative Kompetenz im Sprachgebrauch. Der GER stellt zwar „nur“ eine Empfehlung dar, hat sich jedoch für Lehrwerke, Sprachkurse und bei der Prüfungserstellung als Referenzsystem etabliert. Dies hängt in einem nicht geringen Maße mit der oben beschriebenen übersichtlichen und systematischen Kategorisierung zusammen.

Bei der Prüfungserstellung kommt den Kann-Beschreibungen eine wichtige Funktion zu, da sie in Form von Deskriptoren als Maßstab herangezogen werden: Sie sind als Makrodeskriptoren ausgearbeitet und sollen laut der Autoren des GER je nach Verwendungszweck in Mikrodeskriptoren umgearbeitet werden. In der Praxis werden jedoch die Makrodeskriptoren unmittelbar herangezogen, weshalb die in der Fachliteratur am häufigsten vorgebrachte Kritik am GER darin besteht, dass die Deskriptoren sehr vage formuliert seien und ihre Auslegung sehr großzügig statfinde (vgl.

u. a. Wisniewski 2014; Casper-Hehne 2004; Barkowski 2003). Ein Beispiel für eine vage Formulierung findet sich in Grotjahn/Kleppin (2017), die Begriffe wie „anspruchsvoll“ oder „breites Spektrum“ als nicht eindeutig abgrenzbar bezeichnen. Dieser Sachverhalt wiederum tangiert die Standardisierung von Prüfungen, die einen einheitlichen Maßstab benötigt, der die Grundlage der Bewertung darstellt. Durch die kritisierte vage Formulierung der Deskriptoren wird der Interpretationsspielraum (vgl. Grotjahn/Kleppin 2017) bei der Bewertung vergrößert, was durch eine Formulierung präziserer Deskriptoren relativiert werden kann. Ein Ansatz, die Deskriptoren zu präzisieren, findet sich beispielsweise in Yildirim (2022). In der Untersuchung werden sprachliche Mittel bestimmt, die niveaustufenunterscheidend sind. Zur Präzisierung von Deskriptoren können diese und ähnliche Ergebnisse von Untersuchungen in die Deskriptoren eingearbeitet werden, wodurch eine Erhöhung der Validität von Prüfungen erreicht werden kann.

Die *Validität* gehört neben der *Reliabilität* und der *Objektivität* zu den drei Hauptgütekriterien, deren Einhaltung in der Testforschung als Maß für die Qualität der Prüfung gilt. Die Qualität von Tests ist ein wesentlicher Faktor, der bei Einhaltung der Gütekriterien die Testfairness und -gültigkeit sichert. Bei der *Validität* wird der Frage nachgegangen, ob das, was gemessen werden soll, auch tatsächlich gemessen wird. Am Beispiel der Konstruktvalidität konkretisiert bedeutet dies, dass z. B. Prüfungsaufgaben die entsprechenden sprachlichen Handlungsfelder abdecken (vgl. Demmig 2012, 18). Die *Reliabilität* zeigt an, ob die Messung zuverlässig ist. Das Ergebnis fällt also gleich aus, wenn mehrmals gemessen wird. Beispielsweise kann sich diese Zuverlässigkeit in der Aufgabenstellung zeigen, die so eindeutig gestellt ist, dass die Antwort auch bei mehrfacher Messung jedes Mal die gleiche ist. Bei der *Objektivität* fällt das Ergebnis unabhängig von äußeren Einflüssen oder Bedingungen aus. Allein die Leistung entscheidet über das Prüfungsergebnis, nicht ausschlaggebend ist dagegen z. B. die Einstellung des Bewertenden gegenüber dem Prüfungsteilnehmenden. Um die Gültigkeit von Prüfungen zu gewährleisten, sollten die Hauptgütekriterien möglichst eingehalten werden.

Zur Einhaltung der Gütekriterien ist die Kenntnis testtheoretischer Grundlagen Bedingung. So ist dieses Wissen, wie beispielsweise eine Aufgabe gestellt sein muss, deshalb notwendig, damit die Aufgabe reliabel ist. In dem Personenkreis, der hauptamtlich für die Testerstellung verantwortlich ist, ist die Kenntnis testtheoretischer Grundlagen zur Erstellung von Prüfungen weitestgehend vorhanden. Es gibt jedoch noch weitere Personen, die ebenfalls Tests erstellen, ohne eine eingehende Ausbildung in diesem Bereich erhalten zu haben. Die Prüfungserstellung wird von ihnen oft als „belastendes Anhängsel der Lehre“ betrachtet (Wollersheim/Pengel 2016, 14). Gemeint sind beispielsweise Lehrkräfte, die eine geringe oder keine Beschulung in diesem Bereich haben und Prüfungen ohne oder mit wenig Kenntnis testtheoretischer Grundlage konzipieren. Diese Sachlage wird in der Literatur unter dem Stichwort der fehlenden *Assessment Literacy* diskutiert, indem Herausforderungen im Bereich des Testens benannt werden und eine Auseinandersetzung u. a. mit der Professionalisi-

sierung von Prüfungsentwicklern, die dies nicht hauptberuflich ausüben, stattfindet. Wollersheim/Pengel (2016) schreiben in ihrem Artikel zu *Assessment Literacy* von der „Kunst des Prüfens“, die zur Weiterentwicklung der Qualitätssicherung nicht nur im Bereich des Prüfens, sondern auch im Bereich der Lehre beiträgt.

Ebenfalls die Lehre betrifft der sogenannte *Washback-Effekt*, der im Rahmen des Themas Testen und Prüfen eine zentrale Rolle einnimmt und in diesem Kapitel kurz erwähnt werden soll. Der *Washback-Effekt* ist die Rückkopplung des Unterrichts auf die Prüfung. Der Sprachkurs stellt eine Vorbereitung auf das Prüfungsformat dar, womit eine gezielte Vorbereitung auf die Prüfung erfolgt. „Gemeint sind die Rückwirkungen, die ein Test auf den Unterricht oder auf außerunterrichtliche Testvorbereitungen haben kann, die Kandidaten im Selbststudium vornehmen“ (Demmig 2010, 22). Eine negative Auswirkung für den Prüfungsteilnehmenden kann der *Washback-Effekt* haben, wenn die Prüfung nicht den Kriterien des handlungsorientierten Sprachenunterrichts entspricht und der Sprachenunterricht sich nach dem Prüfungsformat ausrichtet (vgl. Krekeler 2005, 158–160). Nachteilig wird es, wenn beispielsweise die gesamte Prüfung aus geschlossenen Aufgabenformaten besteht und daher auch eine vorrangig rezeptive Vorbereitung erfolgt. Die Handlungsorientierung im Unterricht wird nicht ausreichend berücksichtigt, und die Teilnehmenden erhalten einen einseitigen sprachlichen Zugang. Daher ist die handlungsorientierte Gestaltung von Sprachprüfungen maßgeblich.

3 Prüfungsentwicklung am Beispiel der DSH am Standort Universität Duisburg-Essen

Dieser Abschnitt gibt einen Einblick in die Prüfungsentwicklung, die die Prüfungserstellung, -durchführung und -auswertung umfasst: Das Vorgehen bei der Prüfungsentwicklung und die dabei angewandten qualitätssichernden Maßnahmen werden am Beispiel der DSH am Standort der Universität Duisburg-Essen (UDE) beschrieben. Die Maßnahmen orientieren sich an den Qualitätskriterien der Association of Language Testers in Europe (ALTE).

Vor Beginn der Entwicklungsphase wird eine Beschreibung der Prüfung empfohlen, die relevante Aspekte erfasst (vgl. ALTE, Modul 2, 2–3): Das Prüfungsziel bei der DSH ist das Feststellen der sprachlichen Studierfähigkeit, indem die vier sprachlichen Fertigkeiten Hören, Lesen, Schreiben und Sprechen separat überprüft werden. Die Aufgabentypen der Prüfungsteile, die das Hörverstehen und das Leseverstehen mit Wissenschaftssprachlichen Strukturen überprüfen, sind generell itembasiert, sie sind geschlossen (z. B. MC-Aufgaben), halboffen (z. B. Lückentexte) oder offen (z. B. Kurzantworten auf eine Frage). Des Weiteren beziehen sie sich auf einen (live) vorgetragenen Hörtext bzw. Vortrag (Hörverstehen) und einen Text (Leseverstehen) mit jeweils einer wissenschaftsorientierten Ausrichtung. Beim Schreiben werden die Textproduk-

tionskenntnisse mittels einer vorgabenorientierten Textproduktion evoziert und mithilfe eines Messinstruments bewertet. Der mündliche Ausdruck (Sprechen) wird in einem Gespräch mit zwei Prüfern und mittels eines Sprechanlasses auf Grundlage einer themenbezogenen Vorlage überprüft. Die Prüfungsteilnehmenden sind Studienanwärter, die ihre Hochschulzugangsberechtigung nicht an einer deutschsprachigen Einrichtung erworben haben und in der Regel Deutsch als Fremdsprache gelernt haben. Die Vertrautheit des Testformats ist in den meisten Fällen durch einen absolvierten DSH-Vorbereitungskurs oder C1-Kurs gegeben.

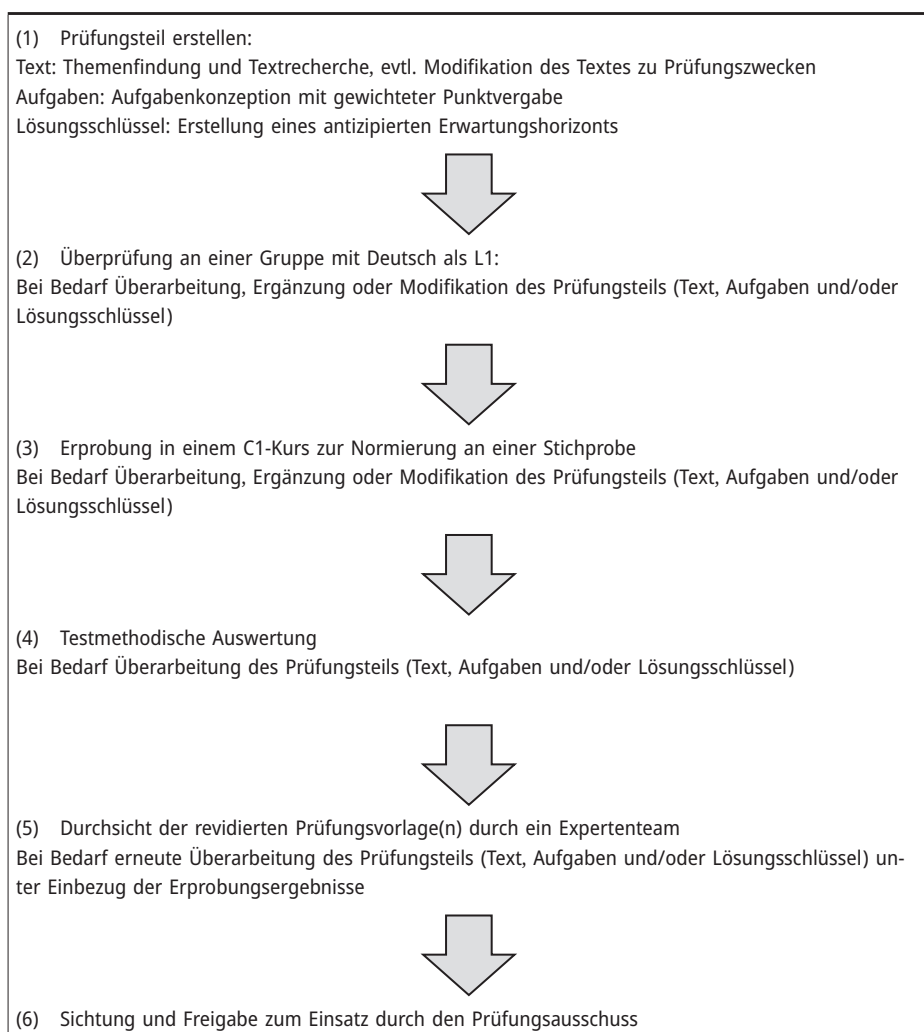


Abb. 1: Phasen der Prüfungsentwicklung.

Zur Erläuterung der Entwicklung von Prüfungsteilen wird exemplarisch die Entwicklung eines Prüfungsteils, der das Leseverstehen überprüft, beschrieben. Die Entwicklung eines Prüfungsteils, der das Hörverstehen überprüft, ist vergleichbar (vgl. Grotjahn/Tesch 2010, 91). Bei unterschiedlichem Vorgehen wird gesondert darauf hingewiesen und mögliche alternative Herangehensweisen werden ergänzt. Das Eingehen auf die rezeptiven Prüfungsteile (Hörverstehen und Leseverstehen) statt der produktiven Prüfungsteile (Textproduktion und Mündlicher Ausdruck) ist der Tatsache geschuldet, dass diese Prüfungsteile bei der Erstellung aufwändiger sind und daher die Beschreibung hierbei detaillierter erfolgen kann. Die Phasen laufen bei allen Prüfungsteilen gleich ab.

Die Erstellung beginnt nach der Themenfindung mit der Textsuche. Dabei ist bei der Auswahl auf bestimmte Merkmale zu achten. Ein Prüfungstext, der die sprachliche Hochschulzugangsbefähigung überprüft, sollte wissenschaftsorientiert und auf C1-Niveau sein. Hierbei sind Kriterien zu beachten, beispielsweise sollen Themen wie Religion oder Krieg ausgeschlossen werden, um keine negativen Emotionen beim Prüfungsteilnehmenden hervorzurufen. (Vgl. ALTE, Modul 3, 4).

Anschließend erfolgt die Konstruktion der Aufgaben. Es gibt mehrere Möglichkeiten, die abhängig von der Teamgröße unterschiedliche Herangehensweisen haben können. Es können sich zur Konstruktion der Fragen mehrere Personen zusammenfinden und gemeinsam abschnittsweise Aufgaben zu dem Text formulieren, die das Leseverstehen überprüfen. Eine andere Herangehensweise ist es, zuerst eine getrennte Erstellung vorzunehmen und dann die Aufgaben gemeinsam zu begutachten oder separat Feedback zu geben. Beim Hörverstehen kann beispielsweise der Text vorgelesen werden, dabei machen sich die Prüfungsentwickler Notizen und konzipieren Aufgaben auf Grundlage ihrer Notizen. Gleichzeitig wird ein Lösungsschlüssel erstellt, das heißt, die erwarteten möglichen Antworten werden aufgeführt.

Nach der Erstellungsphase folgen die Erprobungen der Testteile, um zu überprüfen, ob die Konstruktionen funktionieren:

Um Qualität in Bezug auf Validität und Reliabilität zu gewährleisten, wird zunehmend mehr in die Garantie der statistisch ermittelten Zuverlässigkeit der Ergebnisse investiert. Das heißt: Die in Prüfungen verwendeten Aufgaben werden vor dem Einsatz umfangreichen Erprobungen unterzogen, um ihre Qualität zu garantieren. Erst wenn sich in den Erprobungen gezeigt hat, dass die Aufgaben vom Schwierigkeitsgrad angemessen, unmissverständlich und trennscharf sind, finden sie Aufnahme in die endgültige Fassung der Prüfung. (Perlmann-Balme 2010, 1275).

Zuvor erfolgt eine erste Überprüfung an einer L1-Gruppe, die je nach Zugang aus einer Person bis mehreren Personen bestehen kann. Der Zweck dieser Überprüfung ist es, Konstruktionsschwächen und Verständnisschwierigkeiten aufzudecken. An einer Gruppe ohne sprachliche Schwierigkeiten werden Problematiken in beispielsweise der Aufgabenstellung oder dem Lösungsschlüssel sichtbar. Bei der Überprüfung lösen die Personen die Aufgaben und anschließend werden die Antworten mit dem vorhergesehenen Lösungsschlüssel abgeglichen. Bei Abweichungen zwischen Lösungsschlüssel und den gegebenen Antworten der L1-Gruppe wird festgestellt, ob die Aufgabenstellung unklar ist oder der Lösungsschlüssel Logikfehler aufweist.

Die Erprobung wird unter Prüfungsbedingungen in einer Teilnehmergruppe mit einem vergleichbaren sprachlichen Niveau wie die Zielgruppe durchgeführt. Zu beachten ist, dass bei dieser Gruppe ausgeschlossen ist, dass sie an der Prüfung mit den vorgesehenen Prüfungssets teilnehmen wird. Es muss auch sichergestellt werden, dass sie keinen Kontakt zu Teilnehmenden haben, die an der Prüfung teilnehmen werden, damit keine Informationen weitergegeben werden können. Dies kann beispielsweise durch die örtliche Distanz und die fehlende Verknüpfung zum Prüfungsstandort geschaffen werden.

Die Erprobung wird durch statistische Analysen der Aufgaben ausgewertet. Dabei stehen den Testentwicklern unterschiedliche Verfahren zur Verfügung: einerseits die Methoden der klassischen Testtheorie und andererseits Analyseverfahren aus dem Bereich der probabilistischen Testtheorie (z. B. Rasch-Analyse). Der Einsatz der Rasch-Analyse ermöglicht im Vergleich zu den Methoden der klassischen Testtheorie eine stichprobenunabhängige Auswertung der Daten, setzt aber zum einen fundierte Statistikenkenntnisse voraus und erfordert zum anderen eine größere Probandengruppe (ca. 300 Personen), was nicht alle Testanbieter bewerkstelligen können. Die Item-Analyse nach der klassischen Testtheorie, die wichtige Hinweise zur Qualität der Aufgaben liefert, kann dagegen mit üblichen Kalkulationsprogrammen durchgeführt werden, was sie für einen breiteren Kreis der Testanbieter zugänglich macht. Zu den gängigen Berechnungen zählen die Ermittlung der *Aufgabenschwierigkeit* (wie leicht/schwer die Aufgaben in der Erprobungsgruppe waren), der *Trennschärfe* (wie gut die jeweilige Aufgabe von den leistungsstarken und weniger leistungsstarken Lernern gelöst wurde) sowie die *Distraktoren-Analyse* (wie gut unterschiedliche Antwortmöglichkeiten bei den MC-Aufgaben funktionieren). Die notwendigen Schritte zur Berechnung der genannten Statistiken werden beispielsweise in den Handreichungen für Testautoren der ALTE beschrieben (vgl. ALTE, Modul 2, 16–23.)

Nachdem ein Prüfungsteil erstellt, erprobt, überarbeitet und in einer realen Prüfungssituation eingesetzt wurde, werden die oben aufgeführten statistischen Auswertungen nochmals durchgeführt. Mit den Ergebnissen der Auswertung kann der Prüfungsteil in einem neuen Set nochmals als Prüfungsteil eingesetzt werden, als eine sogenannte Ankerklausur. So können z. B. bei der Erprobung neuer Aufgaben Ankeritems eingesetzt werden, die einen zuvor ermittelten Schwierigkeitsgrad haben, womit die neuen Aufgaben skaliert werden (vgl. dazu z. B. Arras/Grotjahn 2004, 212). Bei Einsatz eines kompletten Prüfungsteils als Ankerklausur wird dieser mit neu erstellten Prüfungsteilen kombiniert, wodurch eine Skalierung des neuen Prüfungsteils erfolgt. Wichtig ist, dass eine bestimmte Zeit vergangen sein muss, bevor die Ankerklausur nochmals eingesetzt wird. Damit wird umgangen oder die Wahrscheinlichkeit sinkt, dass ein Prüfungsteilnehmer, der die Prüfung nicht bestanden hat, den gleichen Prüfungsteil nochmals erhält. In diesem Fall wäre die Prüfungsfairness gegenüber den anderen Prüfungsteilnehmenden nicht gegeben.

Zusätzlich zu den aufgeführten Maßnahmen werden weitere Maßnahmen zur Qualitätssicherung getroffen: Die eingesetzten Bewertenden müssen Erfahrung im Be-

reich DaF haben und werden zudem geschult. Überdies wird bei der Korrektur der Textproduktionen ein analytischer Bewertungsbogen eingesetzt, der eine kleinschrittigere Korrektur der einzelnen Bewertungskriterien erfordert als ein holistischer Bewertungsbogen. Bei der analytischen Bewertung werden die Texte mittels einzelner Kriterien getrennt bewertet. Die Summe der einzelnen Punktwerte führt dann zum Gesamtergebnis. Bei der holistischen Bewertung werden die Texte ganzheitlich über den Gesamteindruck beurteilt. Vor Beginn der Korrektur erfolgt außerdem eine Eichung oder Kalibrierung der Bewertenden. Dieser Vorgang kommt bei der Korrektur von Textproduktionen zum Einsatz, wenn mehr als ein Bewertender beteiligt ist. Bei der Eichung wird zu Beginn ein Text aus der Gesamtmenge der Texte entnommen und für die Anzahl der Bewertenden vervielfacht (beispielsweise gescannt und ausgedruckt oder direkt kopiert). Jeder Bewertende korrigiert und bewertet den Text für sich. Anschließend werden die vergebenen Punkte besprochen. Dabei wird begründet, auf welcher Grundlage die entsprechenden Punkte vergeben wurden. Bei unterschiedlicher Punktezuweisung wird darüber diskutiert und eine Einigung gefunden, indem durch Verallgemeinerung Benchmarks gesetzt werden, an denen sich die Bewertenden orientieren können.

Abschließend lässt sich festhalten, dass aus der Sicht der Teilnehmenden neben der Qualität der Sprachprüfungen auch weitere Faktoren eine wichtige Rolle spielen: z. B. die Höhe der Prüfungsgebühren, passende Prüfungstermine und verfügbare Prüfungsplätze. Wichtig ist in diesem Zusammenhang auch eine maximale Transparenz für die Prüfungsteilnehmenden, beispielsweise durch die Möglichkeit, Einsicht in die jeweiligen Prüfungsordnungen, Durchführungsbestimmungen, Bewertungskriterien und Musteraufgaben zu nehmen. Nicht weniger relevant ist die Forschung zu den bestehenden Prüfungsformaten und ihre wissenschaftlich fundierte Weiterentwicklung bzw. Revision.

4 Prüfungen im Bereich Deutsch als Fach- und Fremdsprache

Da Sprachkenntnisse durch eine bestandene Prüfung nachgewiesen werden müssen, ist es nicht verwunderlich, dass das Angebot an Deutschprüfungen, die im In- und Ausland abgelegt werden können, vielfältig ist. All dies zu behandeln, würde den Rahmen dieses Beitrags deutlich übersteigen. Aus diesem Grund soll der Schwerpunkt der nachfolgenden Ausführungen auf den so genannten Proficiency Tests liegen, die das aktuelle sprachliche Niveau der Teilnehmenden feststellen (s. a. Abschnitt 2 in diesem Beitrag). Eine aktuelle tabellarische Übersicht über die am weitesten verbreiteten standardisierten Proficiency Tests für die Zielsprache Deutsch findet sich z. B. bei Bärenfänger/Geist (2021, 310–311). Dabei wird unterschieden, ob die allgemeine sprachliche Handlungskompetenz geprüft wird oder aber die sprachlichen Kompetenzen innerhalb einer sprachlichen Domäne (vgl. Bärenfänger/Geist 2021, 308).

Im Weiteren wird die mannigfaltige Prüfungslandschaft im Bereich Deutsch als Fremd- und Fachsprache beschrieben. Dabei liegt das Hauptaugenmerk auf dem Zweck der Prüfungen und dementsprechend auch auf deren Zielgruppe. Aus dieser Perspektive betrachtet lassen sich zwei große Gruppen bilden: einerseits allgemeinsprachliche Prüfungen und andererseits zweckorientierte Prüfungen. Die ersteren dienen als Nachweis für ein bestimmtes Sprachniveau und sind in der Regel nach dem GER ausgerichtet. Unter den zweckorientierten Prüfungen werden hingegen Formate zusammengefasst, die sich zwar auch am GER orientieren, aber für bestimmte Zielgruppen entwickelt wurden, so beispielsweise Tests für Zuwanderer, Prüfungen für den Hochschulzugang, berufsbezogene und fachsprachliche Prüfungen. Im Unterschied zu allgemeinsprachlichen Prüfungen wird bei den zweckorientierten Prüfungen oft niveaübergreifend getestet. Zweck und Zielgruppenorientierung beeinflussen konsequenterweise auch den Aufbau der Prüfung und die Themenauswahl.

4.1 Allgemeinsprachliche Prüfungen

Die größten und prominentesten Anbieter der allgemeinsprachlichen Niveauprüfungen für Deutsch als Fremdsprache weltweit sind das Goethe-Institut, das Österreichische Sprachdiplom (ÖSD) und die telc GmbH. Alle drei Institutionen bieten neben Sprachkursen und unabhängig von ihnen Deutschprüfungen auf allen Referenzniveaus des GER (A1–C2) an. Da sich diese Prüfungen auf den GER beziehen, sind sie in Aufbau und Themenauswahl vergleichbar. Die vier Kompetenzbereiche Hören, Lesen, Schreiben und Sprechen sind bei jedem dieser Tests obligatorisch. Bei einzelnen Prüfungen gibt es auch grammatische Aufgaben, die oft in den Prüfungsteil „Lesen“ integriert sind (z. B. telc Deutsch B1 und B2; ÖSD Zertifikat Deutsch Österreich B1).

Neben den Prüfungsformaten für erwachsene Deutsch-Lernende werden auch Prüfungen für Kinder und Jugendliche angeboten, beim Goethe-Institut und telc GmbH auf den Referenzstufen A1 bis B1 und beim ÖSD bis einschließlich C1. Dabei bleiben der Aufbau und der zeitliche Rahmen der Prüfungen gleich, die Themen werden entsprechend den typischen Kommunikationskonstellationen und sprachlichen Handlungen der Kinder und Jugendlichen angepasst (z. B. Schulalltag, Freizeit- und Ferienangebote, Berufswahl).

4.2 Sprachtests für Zuwanderer

Für die Zielgruppe der Zuwanderer, die Deutsch im Rahmen eines Integrationskurses lernen, wurde im Auftrag des Bundesministeriums des Inneren (BMI) vom Goethe-Institut und der telc GmbH gemeinsam der Deutsch-Test für Zuwanderer (DTZ) entwickelt. Der DTZ ist an den GER angelehnt und überprüft sprachliche Kompetenzen auf den Niveaustufen A2–B1 in vier Fertigkeiten: Hören, Lesen, Schreiben und Sprechen.

Neben dem GER dient das Rahmencurriculum für Integrationskurse als Bezugsgröße. So fokussiert der DTZ vor allem sprachliche Domänen und Themen, die für das Leben in Deutschland relevant sind: Wohnen, Arbeitssuche, Aus- und Weiterbildung, Kinderbetreuung, Gesundheit (vgl. z. B. Perlmann-Balme/Plassmann/Zeidler 2009; Plassmann 2017). Zu erwähnen wäre auch, dass der DTZ für Österreich sowohl sprachlich als auch landeswissenschaftlich adaptiert wurde und unter dem Namen Deutsch-Test für Österreich (DTÖ) vom Österreichischen Integrationsfonds (ÖIF) durchgeführt wird.

4.3 Sprachprüfungen für den Hochschulzugang

Wie eingangs erwähnt, ist Deutschland ein attraktiver Studienort für ausländische Studienbewerber. Im DAAD-Bericht aus dem Jahr 2022 wird Deutschland unter Berufung auf die Daten der UNESCO als „das wichtigste nicht englischsprachige Gastland für internationale mobile Studierende“ beschrieben (Pineda et al. 2022, 12). Um ein Studium in einem deutschsprachigen Studiengang aufnehmen zu können, müssen Studienbewerber, die ihre Hochschulzugangsberechtigung im Ausland erlangt haben, in der Regel ihre Deutschkenntnisse nachweisen. Den gesetzlichen Rahmen bildet die „Rahmenordnung über Deutsche Sprachprüfungen für das Studium an deutschen Hochschulen“ (RO-DT), in der die Hochschulrektorenkonferenz (HRK) und die Kultusministerkonferenz (KMK) Prüfungsformate und Qualifikationen festlegen, die als Nachweis der sprachlichen Studierfähigkeit gelten und von den Hochschulen bundesweit anerkannt werden. Unter §2 der RO-DT sind die „Deutsche Sprachprüfung für den Hochschulzugang“ (DSH), der „Test Deutsch als Fremdsprache“ (TestDaF), der Prüfungsteil „Deutsch“ der Feststellungsprüfung (FSP) und das „Deutsche Sprachdiplom der Kultusministerkonferenz“ – zweite Stufe (DSD II) – aufgeführt. Unter befreienden Qualifikationen (§8 RO-DT) finden sich neben den Schulabschlüssen, die der deutschen Hochschulzugangsberechtigung entsprechen, und den Qualifikationen, die ausreichende Deutschkenntnisse nachweisen (z. B. ein abgeschlossenes Germanistikstudium im Ausland), auch die Prüfung „telc Deutsch C1 Hochschule“ und zwei bereits erwähnte allgemeinsprachliche Prüfungen auf dem Niveau C2 des GER, und zwar das „Große Deutsche Sprachdiplom“ des Goethe-Instituts (GDS) und das „Österreichische Sprachdiplom C2“.

In Tabelle 1 finden sich wesentliche Informationen zu einigen dieser Prüfungsformate.

Diese Prüfungsformate zum Nachweis der sprachlichen Studierfähigkeit sind von hoher Relevanz für die Teilnehmenden, weil das Bestehen oder Nicht-Bestehen dieser Prüfungen entsprechend über die Zulassung/Einschreibung zum Studium entscheidet. Mit der Bedeutsamkeit dieser High-Stakes-Prüfungen für die Kandidaten geht auch die Forderung nach der Qualitätssicherung auf allen Etappen des Prüfungsprozesses einher (vgl. z. B. Abschnitt 3 in diesem Beitrag; Wisniewski/Möhring 2021).

Tab. 1: Deutschprüfungen für den Hochschulzugang.

Prüfung	Anbieter/ Durchführungsort	GER-Niveau	Kompetenzbereiche
Deutsche Sprachprüfung für den Hochschulzugang (DSH)	Hochschulen und staatlich anerkannte Studienkollegs in Deutschland	B2–C1	Hören, Lesen, wissenschaftssprachliche Strukturen (Grammatik), Schreiben, Sprechen
Test Deutsch als Fremdsprache (TestDaF)	g.a.s.t. – Gesellschaft für akademische Studienvorbereitung und Testentwicklung e. V.; das TestDaF-Institut/ lizenzierte Testzentren in Deutschland und im Ausland	B2–C1	Hören, Lesen, Schreiben, Sprechen
Prüfungsteil „Deutsch“ der Feststellungsprüfung	Studienkollegs und zuständige Stellen nach Landesrecht	k. A.	orientiert sich an der DSH
Deutsches Sprachdiplom der Kultusministerkonferenz (KMK) – Zweite Stufe (DSD II)	KMK, Zentralstelle für das Auslandsschulwesen (ZfA)/ Regelschulen im Ausland	B2–C1	Hören, Lesen, Schreiben, Sprechen
telc Deutsch C1 Hochschule	telc GmbH/telc Prüfungszentren	C1	Hören, Lesen, Schreiben, Sprechen

4.4 Berufsbezogene Prüfungen

Auch berufsbezogene Sprachprüfungen gewinnen an Bedeutung, was zum einen an der vorschreitenden Globalisierung, der Arbeitsmigration und zum anderen an dem größer werdenden Fachkräftemangel in Deutschland und der Anwerbung internationaler Fachkräfte als Gegenmaßnahme liegt. Fromme/Korb (2018, 50–51) geben einen kurzen Überblick über die aktuellen berufsbezogenen Tests und unterscheiden dabei zwischen einerseits allgemein berufsorientierten Prüfungen (für alle Berufsgruppen) und andererseits fachsprachlichen Prüfungen (z. B. Medizin, Pflege, Recht).

Die nachfolgende Synopse (Tab. 2) gibt Aufschluss hinsichtlich der allgemein berufsbezogenen Prüfungen auf Basis von Fromme/Korb (2018) wie auch auf eigener Recherche, allerdings ohne Anspruch auf Vollständigkeit.

Der Übersicht über allgemein berufsbezogene Prüfungen kann man entnehmen, dass das Angebot im In- und Ausland breit gefächert ist. Gemein ist allen aufgelisteten Formaten, dass es sich dabei um standardisierte Prüfungen handelt, die sich am GER orientieren und sowohl produktive als auch rezeptive Fertigkeiten testen. Die Ausnahme bildet der Goethe-Test PRO, eine computerbasierte und adaptive Prüfung, bei der

Tab. 2: Allgemein berufsbezogene Deutschprüfungen.

Prüfung	Anbieter/Durchführungsort	GER-Niveau	Kompetenzbereiche
Deutsches Sprachdiplom I professional (DSD I PRO)	zFA/ Regelschulen im In- und Aus- land	B1	Hören, Lesen, Schreiben, Sprechen
Goethe-Test PRO: Deutsch für den Beruf	Goethe-Institut/ Goethe-Zentren, Kooperationspartner	A1–C2	Hören, Lesen
ÖSD Zertifikat C2/ Wirtschaftssprache Deutsch	ÖSD/ ÖSD-Prüfungszentren	C2	Hören, Lesen, Schreiben, Sprechen
Prüfung Wirtschaftsdeutsch International (PWD)	DIHK-Gesellschaft für berufli- che Bildung – Organisation zur Förderung der IHK-Weiter- bildung gGmbH/ Auslandshandelskammern; Industrie- und Handelskam- mern; Lizenzpartner	C1	Hören, Lesen, Schreiben, Sprechen
telc-Prüfungen für den Beruf	telc GmbH/ telc Prüfungszentren	A2+, B1+; B1-B2; B2+; B2–C1; C1;	Hören, Lesen (Sprachbausteine), Schreiben, Sprechen
WiDaF-Prüfungen Deutsch als Fremdsprache in der Wirtschaft	Deutsch-französische Indus- trie- und Handelskammer/ akkreditierte Testzentren	A1–B1	WiDaF Basic: Hören, Lesen
			WiDaF Basic Plus: Schreiben, Sprechen
		A2–C2	WiDaF: Hören, Lesen
			WiDaF Plus: Schreiben, Sprechen

nur rezeptive Fertigkeiten überprüft und automatisch ausgewertet werden, so dass die Ergebnisse direkt im Anschluss an die Prüfung abrufbar sind. Adaptiv bedeutet in diesem Kontext, dass alle Teilnehmenden auf dem gleichen Niveau beginnen, die weiteren Aufgaben jedoch auf Grundlage der Lösungen durch das System ausgewählt und an das Niveau angepasst werden (s. dazu auch Klein 2018; Fromme/Korb 2018).

Bezüglich der Funktion und Zielgruppenorientierung erscheint die Prüfungslandschaft ziemlich heterogen: Während Goethe-Test Pro und WiDaF niveauübergreifend testen und somit als Einstufungstest eingesetzt werden können, bietet telc berufsorientierte Prüfungen gezielt für jede Niveau-Stufe zwischen A2+ und C1. Das DSD I Pro (B1) richtet sich an Berufsschüler mit Deutsch als Fremd- und Zweitsprache im In- und Ausland, die Prüfungen PWD (C1) und ÖSD C2 Wirtschaftsdeutsch sind eher für internationale Fachkräfte bestimmt.

Betrachtet man nun das Angebot an Fachsprachprüfungen für Deutsch, so wird deutlich, dass alle im Moment angebotenen Formate Deutschkenntnisse auf dem Gebiet der Gesundheitsversorgung testen (s. Tabelle 3: Fachsprachliche Deutschprüfungen). Diese Tatsache ließe sich einerseits durch den akuten Fachkräftemangel in diesem Bereich und die damit einhergehenden Maßnahmen zur Rekrutierung von ausländischen Ärzten und Pflegekräften erklären. Andererseits spielt die sprachliche Kommunikation beim Medizin- und Pflegepersonal eine entscheidende Rolle, weswegen auch die Anforderungen an die Deutschkenntnisse entsprechend hoch angesetzt sind. Für die Erteilung einer Approbation bzw. einer Berufserlaubnis in akademischen Heilberufen wird das allgemeinsprachliche Niveau B2 sowie das fachsprachliche Niveau C1 vorausgesetzt (vgl. GMK 2014). Im Eckpunktepapier der 87. Gesundheitsministerkonferenz (GMK) wird die Fachsprachprüfung Medizin als Instrument zum Überprüfen der Fachsprachkenntnisse von Ärzten mit Deutsch als L2 festgelegt und die Anforderungen an die Inhalte sowie den zeitlichen Rahmen dieser Prüfung formuliert. Im Fokus stehen dabei die Fertigkeiten Hörverstehen, Schreiben und Sprechen im praxisbezogenen kommunikativen Kontext. Das Fachwissen darf laut dem Eckpunktepapier in keinem Teil der Fachsprachprüfung getestet werden. Eine ausführliche Auseinandersetzung mit der Fachsprachprüfung im Bereich Medizin liefert beispielsweise Jehle (2021). Die Durchführung und Konzeption der medizinischen Fachsprachprüfung obliegt den Ärztekammern der Bundesländer bzw. der Bezirke, die neben der eigenen Fachsprachprüfung auch andere Formate anerkennen dürfen. So werden in einzelnen Bundesländern auch weitere Zertifikate akzeptiert, die sich ebenfalls an den Beschlüssen der 87. GMK orientieren, z. B. telc Deutsch B2–C1 Medizin Fachsprachprüfung und der Patientenkommunikationstest der Freiburg International Academy.

Im Bereich der Pflege sind ausreichende Deutschkenntnisse (in der Regel das GER-Niveau B2) ebenfalls eine Voraussetzung für die Erteilung der Berufserlaubnis. Als Sprachnachweis kann entweder ein allgemeinsprachliches B2-Zertifikat eingereicht oder eine pflegefachsprachliche Prüfung bei anerkannten Aus- und Weiterbildungsstätten abgelegt werden. Auch zwei der bereits erwähnten Testanbieter – das Goethe-Institut und die telc GmbH – haben Prüfungsformate für Pflegeberufe entwickelt (s. auch Klein 2016). Die Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung e. V. sieht in ihrem Positionspapier zu Deutschkenntnissen der Pflegekräfte das Niveau B2 als nicht ausreichend und fordert mit Hinblick auf die sprachliche Realität im Pflegealltag das C1-Niveau als Voraussetzung für die Berufserlaubnis für

Tab. 3: Fachsprachliche Deutschprüfungen.

Prüfung	Anbieter/Durchführungsort	GER-Niveau	Kompetenzbereiche
Fachsprachprüfung Medizin	Ärzttekammern der Bundesländer	C1	Sprechen: Arzt-Patienten-Gespräch und Arzt-Arzt-Gespräch; Schreiben: Kurz-Arztbrief
Patientenkommunikationstest (PKT) Fachsprache C1	Freiburg International Academy	C1	Schreiben: schriftliche Dokumentation Sprechen: Arzt-Patienten-Gespräch und Arzt-Arzt-Gespräch
telc Deutsch Medizin Zugangsprüfung	telc GmbH/ telc Prüfungszentren	B2	Hören, Lesen, Sprachbausteine, Schreiben, Sprechen
telc Deutsch Medizin		B2–C1	Hören, Lesen, Sprachbausteine, Schreiben, Sprechen
telc Deutsch Medizin Fachsprachprüfung		B2–C1	Hören, Lesen und Sprachbausteine, Schreiben, Sprechen
telc Deutsch Pflege	telc GmbH/ telc Prüfungszentren	B1–B2	Hören, Lesen, Sprachbausteine, Sprechen
Goethe Test PRO Pflege: Deutsch für Pflegeberufe	Goethe-Institut, Goethe-Zentren, Kooperationspartner	B2	Hören, Lesen, Schreiben, Sprechen

Pflegepersonen. Die Deutschkenntnisse sollen demnach in einem einheitlichen und standardisierten Verfahren überprüft werden, das neben der allgemeinen sprachlichen Kompetenz auch das Fachvokabular testet (vgl. Thomas/Hauss/Lux 2020).

Neben der Gesundheitsversorgung werden in weiteren Bereichen fachsprachliche Kenntnisse vorausgesetzt. So sind z. B. für die Beeidigung von Dolmetschern bzw. Ermächtigung von Übersetzern zu gerichtlichen Zwecken in Nordrhein-Westfalen das Niveau C2 des GER und sichere Kenntnisse der deutschen Rechtssprache erforderlich. Die entsprechenden Fachsprachprüfungen werden von unterschiedlichen privaten Instituten angeboten. Momentan existieren allerdings keine einheitlichen Vorgaben zum Aufbau und zu den Inhalten dieser Prüfung. Insgesamt betrachtet werden derzeit fachsprachliche Prüfungen nur für wenige Branchen angeboten. Daher stellt die Entwicklung von Testformaten für weitere Berufsgruppen, unter Einbezug aktueller Forschungsergebnisse zur Fachsprache, ein Desiderat dar.

5 Ausblick

Langfristig betrachtet wird sich die Prüfungsentwicklung im Rahmen weitergehender Digitalisierung bewegen. So werden heutzutage neben klassischen papier-basierten Prüfungen auch computer-gestützte (Durchführung im Testzentrum am PC) und reine Online-Prüfungen (ortsunabhängig per Internet) angeboten. Ferner werden derweil unterschiedliche hybride Formate entwickelt, bei denen z. B. der schriftliche Teil in einem Testzentrum und der mündliche Teil online stattfinden kann. Einige der erwähnten Testanbieter haben die Umstellung bereits vorgenommen. So werden beispielsweise der digitale TestDaF, der Goethe Test Pro und die Prüfung Wirtschaftsdeutsch International (PWD) am Computer in lizenzierten Testzentren abgelegt. Zu den hybriden Formaten zählen z. B. die DSH und die meisten telc-Prüfungen, die mündliche Online-Prüfungen zulassen, sowie allgemeinsprachliche Prüfungen des Goethe-Instituts, bei denen der schriftliche Teil am PC und der mündliche Teil ohne Einsatz digitaler Technologien im Testzentrum absolviert wird. Der digitale Wandel betrifft auch die Korrektur bzw. Bewertung der Prüfungsleistungen und eröffnet neue Forschungsperspektiven, z. B. den Einsatz der KI bei Prüfungsentwicklung und -korrektur.

6 Literatur

- Arras, Ulrike/Rüdiger Grotjahn (2004): TestDaF. Jüngste Entwicklungen. In: Hiltraud Casper-Hehne/ Uwe Koreik (Hg.): DSH und TestDaF als hochschulbezogene Prüfungssysteme für Deutsch als Fremdsprache. Baltmannsweiler, 205–225.
- Arras, Ulrike (2007): Wie beurteilen wir Leistung in der Fremdsprache? Strategien und Prozesse bei der Beurteilung schriftlicher Leistungen in der Fremdsprache am Beispiel der Prüfung Test Deutsch als Fremdsprache (TestDaF). Tübingen.
- Association of Language Testers in Europe (ALTE) (2005): ALTE-Handreichungen für Testautoren. (Deutsche Übersetzung von Sibylle Bolton, 2006). Erstellt im Auftrag des Europarats. Frankfurt a. M.
- Auswärtiges Amt (2020): Deutsch als Fremdsprache weltweit. Datenerhebung 2020. 2. Aufl. Berlin.
- Bärenfänger, Olaf/Barbara Geist (2021): Prüfen, Testen, Sprachstände erheben. In: Claus Altmayer/Katrin Biebighäuser/Stefanie Haberzettl/Antje Heine (Hg.): Handbuch Deutsch als Fremd- und Zweitsprache. Kontexte – Themen – Methoden. Berlin/Heidelberg, 301–318.
- Barkowski, Hans (2003): Skalierte Vagheit – der europäische Referenzrahmen für Sprachen und sein Versuch, die sprachliche Kommunikationskompetenz des Menschen für Anliegen des Fremdsprachenunterrichts niveaugerecht zu portionieren. In: Karl-Richard Bausch (Hg.): Der Gemeinsame Europäische Referenzrahmen für Sprachen in der Diskussion: Arbeitspapiere der 22. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen, 22–28.
- Casper-Hehne, Hiltraud (2004): Die Bedeutung des Gemeinsamen Europäischen Referenzrahmens für Unterricht und Prüfungen im Bereich Deutsch als Fremdsprache. In: Hiltraud Casper-Hehne/ Uwe Koreik (Hg.): DSH und TestDaF als hochschulbezogene Prüfungssysteme für Deutsch als Fremdsprache. Baltmannsweiler, 48–60.
- Demmig, Silvia (2010): Backwash Effekt. In: Hans Barkowski/Hans-Jürgen Krumm (Hg.): Fachlexikon Deutsch als Fremd- und Zweitsprache. Tübingen/Basel, 22–23.

- Demmig, Silvia (Hg.) (2012): Sprachkompetenz prüfen – aber wie? Zur Konstruktvalidität von standardisierten Sprachtests am Beispiel der Deutschen Sprachprüfung für den Hochschulzugang (DSH) – testtheoretische Grundlegung und empirische Longitudinalstudie. Frankfurt a. M.
- Fromme, Linda/Eva Korb (2018): Aktuelle Tests und Prüfungen Deutsch für den Beruf. In: *Fremdsprache Deutsch* 59. DOI: <https://doi.org/10.37307/j.2194-1823.2018.59.11>.
- Gemeinsamer europäischer Referenzrahmen für Sprachen (GER): Europarat für kulturelle Zusammenarbeit (Hg.) (2001): Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin, München.
- Glaboniat, Manuela (2010): Sprachprüfungen für Deutsch als Fremdsprache. In: Hans-Jürgen Krumm (Hg.): *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch* 2. Berlin, 1288–1298.
- GMK (2014): 87. Gesundheitsministerkonferenz: Eckpunkte zur Überprüfung der für die Berufsausübung erforderlichen Deutschkenntnisse in den akademischen Heilberufen. https://www.gmkonline.de/documents/TOP73BerichtP_Oeffentl_Bereich.pdf (letzter Zugriff 07. 12. 2023).
- Grotjahn, Rüdiger (2005): Sprachprüfungen als Instrument der Qualitätssicherung in Lehre und Studium. In: Achim Hopbach/Veronique Chalvet (Hg.): *Qualität messen – Qualität managen. Leistungsparameter in der Hochschulentwicklung. Projekt Qualitätssicherung*. Bonn, 55–62.
- Grotjahn, Rüdiger/Karin Kleppin (2017): Gütekriterien bei der Evaluation von Schreibkompetenzen. In: Bettina Akukwe/Rüdiger Grotjahn/Stefan Schipolowski (Hg.): *Schreibkompetenzen in der Fremdsprache. Aufgabengestaltung, kriterienorientierte Bewertung und Feedback*. Tübingen, 41–69.
- Grotjahn, Rüdiger/Bernd Tesch (2010): Messung der Leseverstehenskompetenz im Fach Französisch. In: Bernd Tesch (Hg.): *Standardbasierte Testentwicklung und Leistungsmessung. Französisch in der Sekundarstufe I*. Münster, 91–121.
- Jehle, Nina (2021): Die medizinische Fachsprachenprüfung – Eine Pilotstudie zur Evaluation durch die Teilnehmenden. In: Christina Maria Ersch (Hg.): *Evaluieren und Prüfen in DaF/DaZ*. Berlin, 149–179.
- Kecker, Gabriele (2010): Validität und Validierung von Sprachprüfungen. In: Annette Berndt/Karin Kleppin (Hg.): *Sprachlehrforschung: Theorie und Empirie. Festschrift für Rüdiger Grotjahn*. Frankfurt a. M., 129–146.
- Klein, Wassilios (2016): Ausländische Pflegekräfte: Qualitätssicherung durch zuverlässige Sprachprüfungen. In: *Pflegezeitschrift* 69 (5), 278–281.
- Klein, Wassilios (2018): Didaktik und Methodik der Vermittlung sprachlich kommunikativer Kompetenzen in Aus- und Weiterbildung: Prüfungsformate. In: Christian Efinger/Karl-Hubert Kiefer (Hg.): *Sprache und Kommunikation in der beruflichen Aus- und Weiterbildung. Ein interdisziplinäres Handbuch*. Tübingen, 493–502.
- Krekeler, Christian (2005): *Grammatik und Fachbezug in Sprachtests für den Hochschulzugang*. Essen. URN: <https://nbn-resolving.org/urn:nbn:de:hbz:465-mile-012548-5> (letzter Zugriff 07. 12. 2023).
- Perlmann-Balme, Michaela (2001): Formen und Funktionen von Leistungsmessung und -kontrolle. In: Gerhard Helbig/Lutz Götze/Gert Henrici/Hans-Jürgen Krumm (Hg.): *Deutsch als Fremdsprache. Ein internationales Handbuch*. HSK 19.2. Oldenburg, 994–1006.
- Perlmann-Balme, Michaela (2010): Testen und Prüfen von Sprachkenntnissen. In: Hans-Jürgen Krumm (Hg.): *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch* 2. Berlin, 1272–1288.
- Perlmann-Balme, Michaela/Sibylle Plassmann/Beate Zeidler (2009): *Deutsch-Test für Zuwanderer A2–B1. Prüfungsziele, Testbeschreibung*. 1. Aufl. Berlin.
- Pineda, Jesús/Jan Kercher/Susanne Falk/Theresa Thies/Hüseyin Hilmi Yildirim/Julia Zimmermann (2022): *Internationale Studierende in Deutschland zum Studienerfolg begleiten: Ergebnisse und Handlungsempfehlungen aus dem SeSaBa-Projekt (DAAD-Studien)*. DOI: <https://doi.org/10.46685/DAADStudien.2022.01>.
- Plassmann, Sibylle (2017): Welche Sprache lehren wir? Zur Rolle der Sprachkurse und -prüfungen im Prozess der Integration. In: Annegret Middeke/Annett Eichstaedt/Matthias Jung/Gabriele Kniffka (Hg.): *Wie schaffen wir das?* Göttingen, 325–338.
- Quetz, Jürgen/Raimund Schieß/Ulrike Sköries/Goethe-Institut (Hg.) (2013): *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen*. München.

- Settinieri, Julia/Stefan Jeuk (2019): Einführung in die Sprachdiagnostik. In: Stefan Jeuk/Julia Settinieri (Hg.): Sprachdiagnostik Deutsch als Zweitsprache: Ein Handbuch. Berlin, 3–20.
- Studer, Thomas (2010): Kompetenzmodelle und Bildungsstandards für Deutsch als Fremd- und Deutsch als Zweitsprache. In: Hans-Jürgen Krumm (Hg.): Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch 2. Berlin, 1264–1271.
- Thomas, Vivienne/Armin Hauss/Vera Lux (2020): Sprachkompetenz von ausländischen Pflegefachpersonen – eine Schlüsselqualifikation. Positionspapier der Gesellschaft für Qualitätsmanagement in der Gesundheitsversorgung. URN: https://www.gqmg.de/media/redaktion/Publikationen/Positionspapiere/GQMG_PP_Sprachkompetenz_von_auslaendischen_Pflegefachpersonen_2_Auflage_28.04.20.pdf (letzter Zugriff 07. 12. 2023).
- Wisniewski, Katrin (Hg.) (2014): Die Validität der Skalen des Gemeinsamen Europäischen Referenzrahmens für Sprachen. Frankfurt a. M.
- Wisniewski, Katrin/Jupp Möhring (2021): Die Vergleichbarkeit von Sprachprüfungen zum Hochschulzugang im Licht ausgewählter post-entry-Sprachtests. In: Deutsch als Fremdsprache 4. DOI: <https://doi.org/10.37307/j.2198-2430.2021.04.03>.
- Wollersheim, Heinz-Werner/Norbert Pengel (Hg.) (2016): Von der Kunst des Prüfens – *assessment literacy*. URN: <https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa-212985> (letzter Zugriff 07. 12. 2023).
- Yildirim, Hülya (2022): Ermittlung niveaustufenspezifischer Indikatoren zur Bewertung von Textproduktionen im Rahmen der Deutschen Sprachprüfung für den Hochschulzugang (DSH) – Eine Lernaltersprachenanalyse. Göttingen.