

Standardisierte DaF/DaZ-Tests: Ein notwendiges Übel?

Matthias Jung (Düsseldorf)

1 Historische Situierung der Debatte in DaF/DaZ und Uwe Koreiks Rolle

Eines der Themen, denen Uwe Koreiks Interesse theoretisch wie praktisch in seiner Zeit im FaDaF-Vorstand, als Leiter des Sprachenzentrums der Universität Hannover sowie als Verantwortlicher für die Deutschausbildung an der Türkisch-Deutschen Universität (TDU) galt und gilt, ist das Thema Prüfen und Testen, insbesondere im Zusammenhang mit dem Hochschulzugang und der Studierfähigkeit. Beim FaDaF, der im Auftrag der HRK mit der Fachaufsicht und dem Qualitätsmanagement der DSH betraut ist, hat er sich in den Jahren 2001 bis 2009 intensiv mit der Überarbeitung der DSH-Musterprüfungsordnung (MPO) im Kontext der HRK-Rahmenordnung für den Hochschulzugang (RO-DT) beschäftigt. Anschließend sichtete er zusammen mit Hiltraud Casper-Hehne, der damaligen FaDaF-Vorsitzenden, und der FaDaF-Geschäftsstelle unzählige Prüfungsordnungen der DSH-Hochschulstandorte, ob sie RO-DT bzw. MPO entsprachen, so dass sie offiziell bei der HRK registriert werden konnten. Denn erst die Registrierung garantierte, dass die jeweilige DSH sich auch als solche bezeichnen durfte und deutschlandweit als Sprachnachweis für den Hochschulzugang anerkannt wurde. Von Anfang an ging es dabei neben diesen Formalia auch um die Frage der Qualität und Vergleichbarkeit der verschiedenen Sprachprüfungen für den Hochschulzugang.

Diese Debatte wurde insbesondere durch die Einführung des TestDaF forciert, der 2001 zum ersten Mal weltweit abgenommen wurde und mit dem ein neuer wissenschaftlicher Anspruch im Bereich der DaF-Tests Einzug hielt. Vor allem die DaF-Lehrgebiete, aber auch der damalige FaDaF-Vorstand begleiteten die Entwicklung und Einführung des TestDaF vielfach sehr kritisch und hätten seine Einführung wohl gerne verhindert, weil sie den weltweit einheitlichen Hochschulzugangstest als Konkurrenz zu der von den Hochschulen und das hieß seinerzeit fast immer von den DaF-Lehrgebieten dezentral verantworteten DSH sahen. Die damals in manchen Sitzungen aufwallenden Emotionen und die damit einhergehende manchmal scharfe persönliche Polemik sind ein gutes Beispiel dafür, dass es bei standardisierten Tests bereits für die Prüfungsanbieter offensichtlich um sehr viel geht, und es sich nicht um eine rein wissenschaftliche Angelegenheit handelt. Das gilt erst recht für die Prüflinge, die ganz andere Prioritäten als testtheoretische Diskussionen haben (siehe Abschnitt 4).

Mit dem Wechsel praktisch des gesamten FaDaF-Vorstands 2001 versachlichte sich die Debatte, was nicht zuletzt ein Verdienst des neuen FaDaF-Vorstandsmitglieds und damaligen FaDaF-Geschäftsführers Uwe Koreik war. Denn obwohl er sich parallel sehr engagiert für die Überarbeitung und Modernisierung der DSH einsetzte, ging es ihm nicht um die Polemik, sondern um den wissenschaftlichen Diskurs über Testen und Prüfen im Fach DaF (vgl. Casper-Hehne; Koreik 2004, Koreik 2005).

2 Der DaF/DaZ-Test-Boom

Seit dieser Zeit, d.h. den späten 1990er und frühen 2000er-Jahren hat sich zum einen das Fach DaF zu einem Fach DaF/DaZ weiterentwickelt, zum anderen sind die Relevanz des Themas Testen & Prüfen ebenso wie die Menge der unterschiedlichen standardisierten und deutschlandweit anerkannten DaF/DaZ-Prüfungen und die Zahl der abgelegten Tests stark gestiegen. Im DaZ-Bereich sind besonders die B1-Abschlussprüfungen der Integrationskurse für Migrant:innen wie der DTZ (Deutschtest für Zuwanderer), aber auch die B2- und C1-Prüfungen zum Abschluss der sogenannten Berufssprachkurse zu nennen, die seit 2022 durch den neu entwickelten, dann einheitlich vorgeschriebenen DTB (Deutschtest für den Beruf) abgelöst werden. Und auch im Bereich des Hochschulzugangs hat sich ein weiterer Test eines Anbieters etabliert, der im BAMF-Bereich groß geworden ist: die telc GmbH, hinter der als dominanter Anteilseigner der Deutsche Volkshochschulverband (DVV) steht. Hinzu kommen im Kontext der Fachkräftegewinnung insbesondere im Bereich Medizin & Pflege weitere berufsspezifische Tests, die einen Zugang zu den reglementierten Berufen wie Ärzt:in oder Pflegefachkraft ermöglichen.

An das Bestehen oder Nicht-Bestehen dieser Tests sind weitreichende gesellschaftliche und ökonomische Konsequenzen geknüpft. DaF/DaZ-Prüfungen entscheiden über Aufstiegs- und Lebenschancen und das oft in einem ungleich drama-

tischeren Ausmaß als Abiturprüfungen oder ein Hochschulexamen, denn von ihrem Bestehen hängen z.B. Visaerteilung, Bleiberecht und Abschiebung, Staatsbürgerschaft, Zulassung zum Studium, Arbeitserlaubnis, Berufszulassung, Arbeitsverträge, Familienzusammenführung, Sozialhilfen u.v.a.m. ab.

Damit einher ging nicht nur die Entwicklung der entsprechenden wissenschaftlichen Disziplin im Bereich der Sprachtests, sondern auch der starke Glaube an die Messbarkeit von kommunikativer Kompetenz und Objektivität der Ergebnisse, was vielfach, und hier ist vor allem der Bereich der BAMF-Kurse zu nennen, zu einem technokratischen Selbstläufer wurde. Dabei spielte noch eine weitere Entwicklung der letzten 20 Jahre eine wichtige Rolle: Der 2001 eingeführte Gemeinsame Europäische Referenzrahmen für Sprachen (GER) hat zu einer eindeutigen Strukturierung und formalen Vergleichbarkeit des Prüfungsangebots geführt bzw. die Illusion der Äquivalenz von Prüfungen auf derselben GER-„Stufe“.

3 Die Testproblematik als nicht-wissenschaftliche Frage

Angesichts der ständig wachsenden Bedeutung von Deutschzertifikaten für individuelle Lebenswege scheint es an der Zeit, den damit einhergehenden blinden Glauben an standardisierte Tests in Frage zu stellen und sie als das zu sehen, was sie sind: ein vielfach notwendiges Übel. Anstelle einer Verengung der DaF/DaZ-Perspektive auf Fragen der wissenschaftlichen Methodik und der Testkonstruktion soll der Blick auf den Prozess der Chancenzuweisung durch Testergebnisse und ihre Instrumentalisierung gelenkt werden. Die meisten dieser Aspekte sind nicht neu, aber sie müssen in Erinnerung gerufen und im Licht der Entwicklungen der letzten Jahre speziell für DaF/DaZ neu bewertet werden.

Die prinzipielle Kritik an standardisierte Tests konzentriert sich auf ihren Charakter als summative Prüfungen, die das Erreichen eines vorgegebenen Lernstands bestätigen sollen, während die pädagogisch „guten“ Prüfungen formativer Art sind, d.h. als individualisiertes Feedback den Fortschritt der Lernenden steuern und sie motivieren.

Die pädagogische Grundsatzkritik an summativen Prüfungen, speziell wenn sie bestimmte, lernendenunabhängige Standards vorgeben, ist nicht neu: Summative Tests machen weder das Potenzial von Lernenden sichtbar noch sind sie individuell fair in dem Sinne, dass sie die gemachten Fortschritte oder unterschiedliche Lernbedingungen berücksichtigen. Hinzu kommen generelle Vorbehalte gegen den Leistungsdruck, der Versagensängste provoziert und die Lernenden demotiviert.

Aus Sicht der Lehrenden wird beklagt, dass die Fixierung auf die Prüfungen bei den Lernenden wie der entsprechende Druck auf die Lehrkräfte, möglichst hohe Bestehensquoten in ihren Klassen zu erreichen, den Unterricht auf eine reine Prüfungsvorbereitung reduziert („Washback-Effekt“), bei der nicht abgeprüfte, meist komplexere Lernziele, insbesondere Fertigkeiten bzw. Verhalten und Werte im

sozialen und kommunikativen Bereich zugunsten einer Wissensüberprüfung und operationalisierbaren Teilkompetenzen entfallen.

Auf diese Kritik hat speziell die Hochschullehre mit dem Konzept des „Constructive Alignments“ (Biggs; Tang 2011) reagiert. Gemeint ist, vereinfacht gesagt, dass Lernziele nicht heimlich durch die Prüfung statt durch das Curriculum vorgegeben werden, sondern umgekehrt: Leistungstests sollen so konstruiert werden, dass Lernzielerreichung und Prüfungsvorbereitung aufgrund ihrer gleichen „Ausrichtung“ zusammenfallen. Das ist immer dann besonders gut möglich, wenn die Lehrkräfte auch die Prüfung selber entwickeln bzw. beeinflussen können. Ein *constructive alignment* wird aber umso schwieriger, je standardisierter und weiter verbreitet eine Prüfung ist, so dass sie generell nur noch unter großem Aufwand und vielleicht alle zehn oder zwanzig Jahre verändert und erst recht nicht auf individuelle Lernziele angepasst werden kann.

Weitere Stoßrichtungen der Kritik stellen die Objektivität und Vergleichbarkeit der Bewertungen („Reliabilität“) bzw. den Nachweis des Erreichens eines Lernziels bzw. den Erwerb einer konkreten Kompetenz durch eine bestimmte Prüfung („Validität“) in Frage.¹ Dabei gerieten vor allem verschiedene Formen sachfremder Verzerrungen („bias“) in den Blick, seien es soziale Vorurteile der Bewertenden oder kulturelle, z.B. eurozentrische Prüfungsaufgaben.

Diese Kritik an Prüfungen im Allgemeinen trifft auch auf Sprachprüfungen zu, wenn auch in jeweils eigener Ausformung und Gewichtung, speziell auch in der Kombination mit dem Gemeinsamen Europäischen Referenzrahmen für Sprachen (GER). Hier gibt es zusätzliche Kritikpunkte, etwa die Einteilung in „Niveaustufen“, als ginge es um wohldefinierte Plateaus einer Treppe, wo es sich doch in Wirklichkeit um ein Kompetenzkontinuum handelt, in dem relativ willkürlich Schnitte gesetzt werden, die dann als Grenze zwischen den sechs definierten Niveaus A1 bis C2 mit fließenden Übergängen gelten (und wo Lehrbuchverlage wie Sprachkursorganisatoren noch willkürlicher meist Halb- und Übergangsstufen wie A1.1/A1.2, B1.1 und B1.2. oder B1+ implizieren).²

Schon diese Bildlichkeit ist also falsch, denn die individuelle Sprachkompetenz entwickelt sich nicht in wohldefinierten Sprüngen, sondern als kontinuierlicher Prozess mit Fort- und Rückschritten und das oft gleichzeitig. Leider hatte die GER-Standardisierung verschiedener Sprachkompetenzniveaus die Konsequenz, dass mittlerweile gerade auch im institutionellen Bereich mit den GER-Abkürzungen von A1 bis C2 operiert wird, als handele es sich um eindeutige technische Normen, Maße und Gewichte oder diskrete Führerscheinklassen, mit denen sich die Kommunikationskompetenz eines Menschen präzise klassifizieren ließe. Diese Kritik

¹ Ein schönes Beispiel für eine nicht-valide Prüfungsfrage gibt Juli Zeh (2012) in ihrer Erzählung „Nullzeit“, in der ein Jura-Student in seiner mündlichen Staatsexamensprüfung nicht nur den Namen des Rechtsphilosophen Montesquieu buchstabieren soll, sondern zusätzlich noch nach dem vollständigen Namen des Philosophen inkl. aller seiner Titel gefragt wird.

² Diesen Punkt führt etwa Claudia Harsch genauer in ihrem Online Vortrag unter <https://www.youtube.com/watch?v=XQUwxPhd5kc> aus. Für weitere Kritikpunkte am GER s. DGFF (2018).

betrifft allerdings nicht das Konzept des GER selber, sondern seine Anwendung in der Praxis vor allem durch Fachfremde.³

Im Folgenden soll es allerdings nur am Rande um die genannten und bekannten Kritikpunkte gehen, sondern entweder um „unfachliche“ Perspektiven oder um prinzipielle Fragen, ob – speziell im Bereich der Integration von Geflüchteten oder Fachkräften – die Fixierung auf standardisierte Kompetenzmessungen der einzige und vor allem der beste Weg zum Ziel ist.

4 Die Perspektive der Prüfungsteilnehmenden⁴

Ganz und gar unwissenschaftlich ist z.B. das Kriterium der „Schwere“ einer Prüfung (verstanden als die subjektive Erwartungswahrscheinlichkeit durchzufallen) oder die organisatorischen Rahmenbedingungen der Prüfungsdurchführung (Verfügbarkeit von Prüfungsplätzen, Ergebnisfristen), die einen großen Einfluss auf die Wahl der einen oder anderen Hochschulzugangsprüfung durch ausländische Studienbewerber:innen haben. Diese Faktoren beeinflussen die Selektion und Zahl der ausländischen Studierenden in Deutschland in ungleich höherem Maße als die wissenschaftliche Qualität einer Prüfung, die dagegen bei der Entscheidung über die Anerkennung einer Prüfung für einen bestimmten Zweck wie den Hochschulzugang eine Rolle spielt, aber auch hier nur ein Faktor unter mehreren ist.

Noch grundsätzlicher ist die Frage, ob die standardisierten Prüfungen, in unserem Kontext etwa Hochschulzugangsprüfungen wie die DSH oder der TestDaF oder die Abschlussprüfungen der BAMF-Kurse, überhaupt wünschenswert oder zumindest notwendig sind. Denn jenseits des gesteuerten Sprachunterrichts gibt es nicht nur Kinder, die eine Sprache weitgehend selber aneignen, sondern auch *Erwachsene*, sei es weil sie durch aktive sprachliche Interaktion mit ihrer zielsprachlichen Umgebung besonders gut eine neue Sprache auf „natürlichem“ Weg quasi von selbst lernen, sei es weil sie gute autonome Lernende sind, d.h. selbst gesteuert und reflektiert mit Sprachlernmaterialien ihren Spracherwerb organisieren oder dies in einer Kombination beider Verfahren tun.

Auch wenn ich dafür nur anekdotische Belege aus vielen Gesprächen kenne, scheinen z.B. Fälle, in denen Bewerber:innen aus dem Ausland sich mit nur geringen Deutschkenntnissen regulär immatrikulieren konnten und anschließend nicht nur eine hohe Sprachkompetenz erwerben, sondern auch ihr Studium erfolgreich abschlossen, gar nicht so selten. Solche Fälle lassen sich mit dem oben aufgeworfenen Punkt erklären, dass es Personen mit hoher Deutscherwerbseignung und -motivation gibt, die in ihrem Heimatland keinen Zugang zu den notwendigen Deutsch-

³ Zur Anwendung des GER für Tests vgl. ALTE (2012) und British Council, UKALTA, EALTA, ALTE (2022).

⁴ Um Fairness von Prüfungen und ethische Fragen geht es auch im Handbuch für Testersteller, wenn auch nur kurz (ALTE 2012: 21f.).

kursen bzw. zu ungesteuerten Spracherwerbssituationen hatten, den sie dagegen in der zielsprachlichen Umgebung problemlos finden.

Angesichts der immer lückenloseren und professioneller angewandten Sprachregularien für Einschreibungen und Visa sind die verbliebenen Schlupflöcher, um der standardisierten Sprachkompetenztestung zu entgehen, betrügerischer Art: Zeugnisfälschungen, außerdem „echte falsche“ Zeugnisse, bei denen ein:e Stellvertreter:in die Prüfung bestanden hat, sowie der Betrug durch den vorherigen Zugang zu den Prüfungsaufgaben. Es liegt im Wesen dieser in den Prüfungsordnungen sanktionierten Verstöße, dass die Dunkelziffer hoch ist.

Verkäufer:innen falscher Zertifikate locken ihre Kundschaft nicht zuletzt mit dem Argument, dass die Tests sowieso nicht realistisch seien. In einem entsprechenden Spameintrag in einem offenen Forum heißt es etwa unter der entwaffnenden Überschrift „Buy Language Certificates“ und in etwas schrägem, wohl aus dem Englischen schlecht übersetztem Deutsch:

Lesen Sie weiter, um herauszufinden, was ein Sprachzertifikat für Sie tun kann und wie Sie es erhalten. Warum ein Sprachzertifikat anstreben? Lassen Sie Ihre Fähigkeiten ehrlich von einem Dritten bewerten [...]. Haben Sie eine schnelle Antwort auf Fragen zu Ihren Sprachkenntnissen parat. Für Beschäftigung, Studium oder Einwanderung berechtigt sein 5 der nützlichsten Sprachzertifikate der Welt JLPT (Japanisch) DELE (Spanisch) DELF/DALF (Französisch) HSK (Mandarin-Chinesisch) TELC (mehrere Sprachen). So sehr die Designer des Tests auch versuchen, den Sprachgebrauch in der realen Welt widerzuspiegeln, wird er für die Benotung zwangsläufig ein wenig künstlich sein. Daher garantiert ein hohes Niveau in Ihrer Zielsprache niemals eine solide Punktzahl.⁵

Wer solche unlauteren Wege geht, tut dies, soweit er z.B. überhaupt vorhat ernsthaft zu studieren, in der festen Überzeugung, dass er auch ohne den vorherigen Nachweis der als notwendig erachtete Kompetenzstufe sprachlich schon irgendwie klar kommt bzw. seine sprachlichen Defizite mit Übersetzungsprogrammen oder mit Hilfe seiner Community und weiteren freundlich gesagt „Tricksereien“ kompensiert. Aber auch die Haltung, dass wenn man erst einmal im Land ist, die Sprache schon irgendwie lernt, dürfte verbreitet sein. Und obwohl sich die Mogler:innen in den allermeisten Fällen täuschen, gibt es auch Ausnahmen, denen es tatsächlich gelingt, ein Studium erfolgreich auf Deutsch zu absolvieren.

⁵ Dieser SPAM tauchte am 18.7.2022 in allen Diskussionsforen auf deutsch-als-fremdsprache.de auf und wurde natürlich umgehend gelöscht bzw. gar nicht erst freigeschaltet.

5 Legale Strategien zur Verfälschung der Testvalidität

Neben den genannten, eindeutig betrügerischen Strategien, um an einen notwendigen Sprachnachweis zu gelangen, gibt es aber auch Grauzonen, die den Glauben an die Objektivität eines standardisierten Sprachtests ins Wanken bringen. Das gilt vor allem für die Schreibaufgaben und allgemein im Hinblick auf die notwendigerweise begrenzte Zahl von Prüfungssätzen der Testinstitute. Hier lassen sich im Netz dann Themenlisten aller bisherigen Schreibaufgaben finden, für die man natürlich entsprechende Texte vorbereiten kann, um sich dadurch einen objektiven Vorteil zu verschaffen. Überhaupt gibt es in einer Art Prüfungsdarknet zahlreiche Hinweise zu den aktuell eingesetzten Prüfungssätzen, die auf verschiedene Weisen geleakt wurden, sei es durch korrupte Prüfer:innen oder Testinstitute, durch Abfotografieren mit dem Handy, Entwenden von Prüfungssätzen oder seit der Einführung digitaler Tests natürlich auch durch Hacking bzw. Phishingmethoden. Dass dahinter oft eine gehörige Portion krimineller Energie steckt, vor allem wenn diese Informationen dann monetarisiert werden, ist unbestritten. Die Testinstitute sind hier in einem ständigen Hase-Igel-Wettlauf gefangen, mit dem sie diese Verfälschungen zwar einigermaßen begrenzen, aber niemals vollkommen unterbinden können, auch nicht technisch. Was macht man z.B. gegen das genaue Memorieren von Prüfungssätzen bzw. seinen Teilen, erst recht, wenn sich mehrere Personen zusammentun und auf einzelne Teile konzentrieren, vielleicht sogar nur zu diesem Zweck an einer Prüfung teilnehmen?

Unabhängig von der kriminellen Energie der Täter:innen ist es nicht verboten, Listen mit typischen Themen o.ä. Prüfungshinweisen zu benutzen, ja es gehört sogar zu einer seriösen Vorbereitung auf eine Prüfung sich möglichst viele Informationen über diese Musterprüfungssätze etc. zu besorgen, so dass der Übergang zur vollkommen legalen Prüfungsvorbereitung fließend ist. Reine Prüfungsvorbereitungen, die sogenannten Formattrainings, wollen keinen Kompetenzzuwachs im Sinne des GER erzielen, sondern nur mit einer Prüfung so vertraut machen, dass man seine volle Sprachkompetenz abrufen kann und nicht wegen mangelnder Vertrautheit mit den Aufgabentypen und -bedingungen einer Prüfung aus dem Konzept gebracht wird und ein Ergebnis unter seinem wirklichen Sprachniveau erzielt. Soweit die Theorie eines guten Formattrainings.

Vergessen wir einmal, dass die Verfügbarkeit bzw. Auswahl von Prüfungstrainings, aber auch das finanzielle Potenzial der Kandidat:innen, was die Bezahlung solcher Kurse und mehrfach wiederholter Tests angeht, eine grundsätzliche Chancenungleichheit schafft, dann gibt es noch die Vorbereitungstrainings, die den Glauben an die Objektivität von Tests erschüttern. Gemeint sind Formatdrills, die so perfektioniert auf einen bestimmten Prüfungstyp vorbereiten, dass dieser bestanden wird, *ohne* dass man das im Zeugnis anschließend nachgewiesene Niveau hat, man also Testergebnisse über seiner realen Sprachkompetenz erzielt. Die bei diesen Drills eingesetzten Mittel können einerseits die Nutzung aller genannten, auch „grauer“ Informationsmöglichkeiten über aktuelle Prüfungssätze, das Auswendig-

lernen vorbereiteter Texte und Versatzstücke sein und wie man sie clever kombiniert, das Wissen über Prüfungsstrategien (z.B. auch dann bei Multiple-Choice-Tests etwas ankreuzen, wenn man die Antwort nicht weiß und in diesem Fall gerade nicht die verführerischen Distraktoren ankreuzen, die ein Wort aus dem Hör- oder Lesetext zitieren). Solche im Sinne der zu messenden Kompetenz nicht-validen Faktoren können über Bestehen oder Durchfallen entscheiden. Bestenfalls lässt sich argumentieren, dass eine gründliche Prüfungsvorbereitung zwar eine nicht-sprachbezogene, aber eben eine wichtige Studier- und Karrierekompetenz abbildet.

Wie weit es entsprechenden Kursen tatsächlich gelingt, ohne direkt unlautere Mittel, wie ein vorher bekannter Prüfungssatz, den objektiven Kompetenzmessungsanspruch eines Tests auszuhebeln, sehen Sprachtestinstitutionen und Sprachlehrkräfte naturgemäß sehr anders. Letztere betonen, wie sie ihre Lernenden durch Prüfungsdrills oder Strategien wie vorbereitete Textproduktionen so fit gemacht haben, dass auch schwache Kursteilnehmende bestanden haben. Oder sie zitieren erlebte Fälle, in denen ihre schlechtesten Lerner:innen bestehen und die besten durchfallen oder in denen Studienbewerber mit anerkannten Hochschulzugangszugnisse in Wirklichkeit eine Deutschkompetenz weit unter der zertifizierten hatten, nichts verstanden oder selbst einfachste Gespräche unmöglich waren. Wie weit solche anekdotischen Erzählungen, die es zuhauf gibt, als Evidenz für die mangelnde Validität einer bestimmten Prüfung taugen – zumindest in dem Sinn, dass dies mit bestimmten Tricks möglich ist – bleibt offen, nicht zuletzt, weil die geschilderten Phänomene auch anders, und nicht nur durch Betrug, erklärt werden können.

Zudem verweisen die Prüfungsanbieter auf ihre Maßnahmen, um genau solche Fälle zu vermeiden (sie aber realistischerweise dadurch aber wohl nur mehr oder weniger erschweren). Als Strategie zur Aufrechterhaltung der Prüfungsvalidität taugt es beispielsweise im Testkonstrukt sich für viele kleine Schreibaufgaben zu entscheiden, die in den Prüfungssätzen immer wieder anders kombiniert werden können, während eine einzige längere Schreibaufgabe zu aktuell gesellschaftsrelevanten Themen in der Tat oft gut vorhergesehen und mehr oder weniger durch Auswendiglernen vorbereitet werden kann. Auch ständig neue Prüfungssätze konkreter unläutere Prüfungsvorbereitungen, sind aber aufwendig und teuer, so dass kein Anbieter von standardisierten Prüfungen es sich erlauben kann, Prüfungsteile nur einmal einzusetzen.

Schließlich spielen auch die Richtlinien für die Bewerter:innen eine Rolle bei der Sanktionierung von Betrugsversuchen, z.B. wenn offensichtlich vorher gelernte Textstücke oder gute Texte, die aber das Thema verfehlen, abgestraft werden. Bewertungsvorgaben, die beispielsweise auch festlegen, was als Täuschungsversuch automatisch zum Nichtbestehen der gesamten Prüfung führt, werden allerdings nicht veröffentlicht und nur bei begründeten Einsprüchen manchmal offengelegt. So hat in einem mir bekannten Fall der Prüfungsanbieter telc eine gesamte C1-Prüfung als Täuschungsversuch und damit als nicht bestanden gewertet, weil das Ergebnis des Testteils „Schreiben“ von den anderen Teilnoten stark nach unten

abwich. Dabei hatte der Kandidat trotz der deutlich geringeren Punktzahl in der Fertigkeit Schreiben nach der Prüfungsordnung aufgrund seiner Gesamtpunktzahl bestanden.

In diesem konkreten Fall waren sich die beiden Dozentinnen des betreffenden Lernalters in dem C1-Berufssprachkurs sicher, dass hier kein Betrugsversuch vorlag: Sie hatten für diesen Lerner auf der Basis des viermonatigen gemeinsamen Kurses das Bestehen der C1-Prüfung erwartet und wussten um seine relative Schreibschwäche, er hatte zudem erklärt für diesen Teil in der Prüfung zu spät gekommen und daher besonders nervös gewesen zu sein. Warum sollte er in den anderen Teilen erfolgreich betrogen haben, dann im schriftlichen Teil aber nicht? Ist es unplausibel, dass man in einer Fertigkeit/einem Testteil im Prüfungsstress (relativ) versagt? Umgekehrt, d.h. bei drei konsistent schwachen Ergebnissen in allen anderen Prüfungsteilen und einer brillanten Textproduktion liegt der Betrugsverdacht viel näher. Es geht hier nicht um den Einzelfall, der nur als Beispiel dienen soll, dass das notwendige Vorgehen gegen Prüfungsbetrug wieder zu anderen Ungerechtigkeiten und Verfälschungen führen kann.

6 Fertigkeitenmix und Gültigkeit von Sprachzertifikaten

Dieser Fall bringt uns zu weiteren Kritikpunkten hinsichtlich der Validität verschiedener Sprachtests jenseits der testtheoretischen Güte der einzelnen Aufgaben und aller Sicherheitsvorkehrungen gegen Täuschungsversuche: Die Regeln für die Wertung von Teilresultaten im Rahmen des Gesamtergebnisses betreffen das Testkonstrukt insgesamt. Teilergebnisse differenzieren meist nach den vier Fertigkeiten Hören, Sprechen, Lesen, Schreiben, wobei manchmal noch andere Teilnotenbereiche hinzukommen. Doch wie werden diese Ergebnisse für das Bestehen der Gesamtprüfung zusammengezählt?

Die laut Visumshandbuch⁶ anerkannten Sprachnachweise für die wichtige B1-Stufe, insbesondere relevant nach dem Fachkräfteeinwanderungsgesetz, das B1-Zertifikat des Goethe-Instituts (GI) und das Österreichische Sprachdiplom (ÖSD) auf der einen und das B1-Zertifikat der telc GmbH auf der anderen Seite, unterscheiden sich inhaltlich nur wenig, der größte Wertungsunterschied liegt in den Kumulationsregeln. Während für die ersten beiden Zertifikate alle vier Einzelmodule bestanden sein müssen, um das B1-Zertifikat zu erhalten, müssen beim telc-Test jeweils nur das Bestehen der drei schriftlich abgefragten Teile insgesamt und die mündliche Prüfung bestanden sein, d.h. ein schlechter HV-Test, kann durch einen überdurchschnittlichen LV und eine gute Schreibleistung kompensiert werden.

Es lässt sich lang darüber diskutieren, welches Testkonstrukt sinnvoller bzw. fairer ist. Entscheidend ist hier für mich allerdings der Punkt, dass das modulbezogene Bestehensprinzip von GI und ÖSD mehr Chancen zum Durchfallen bietet als

⁶ Im Visumshandbuch (Auswärtiges Amt 2022) wird an 25 Stellen auf das Niveau B1 Bezug genommen, die anerkannten Zertifikatsanbieter werden auf S. 6 des Abschnitts „Nachweis von Sprachkenntnissen im Visumsverfahren“ genannt.

das kumulative telc-Prinzip. Leider geben sich die Prüfungsanbieter bei den Statistiken zu ihren Prüfungen aus verschiedenen Gründen wenig transparent – am ehesten übrigens noch g.a.s.t. als verantwortlicher Anbieter für den TestDaF –, so dass ich hier nur mit einem kleinen Sample von knapp 150 Prüfungskandidat:innen operieren kann, für das mir die Zahlen bekannt sind und die das ÖSD in Tunesien abgelegt haben: Wäre das kumulative Wertungsprinzip wie bei telc angewendet worden, hätten rund 10% mehr der Kandidat:innen im ersten Versuch bestanden. Diese Zahl ist gruppenabhängig, aber es können in keinem Fall mehr Personen nach dem Einzelmodulprinzip als nach der schriftlich kumulativen Wertung bestehen. Auch hier liegt eine weitere Ergebniszufälligkeit infolge des gewählten Tests vor, geht man einmal davon aus, dass die beiden Testfamilien aufgrund ihrer Qualitätsgarantie durch die ALTE-Zertifizierung in etwa gleich valide sind.

Zu hinterfragen ist allerdings grundsätzlicher, ob ein Sprachtest, mit dem man nachweisen möchte, dass man einer angestrebten Tätigkeit wie studieren oder in einem bestimmten Beruf arbeiten auch tatsächlich kommunikativ gewachsen ist, immer alle vier Fertigkeiten auf gleichem GER-Niveau erfordert. Wer sich um ein Au-pair-Visum bemüht, muss vermutlich auf einem höheren Niveau sprechen und ein über ein gut entwickeltes Hörverständnis verfügen als lesen und schreiben können. Der Erkenntnis, dass nach dem GER-Ansatz eigentlich die einzelnen Fertigkeiten je nach kommunikativen Erfordernissen der angestrebten Tätigkeit differenziert werden müssten, trägt zwar der TestDaF Rechnung, indem er das individuelle Testergebnis nach TestDaF-Kompetenzstufen (TDN) differenziert, so dass auch für einzelne Studiengänge unterschiedliche Fertigungsprofile vorgeschrieben werden könnten. Doch wendet das m.W. leider keine deutsche Hochschule an, sondern sie greifen zur Differenzierung der Bewerber:innen auf ein kumuliertes „Punkte“-System von 12 (4 x TDN 3) bis 20 (4 x TDN 5) Punkten zurück. Hier zeigt sich erneut, dass bei der Nutzung von Sprachnachweisen in Verwaltungsvorgängen, die über Lebenswege entscheiden, fachfremde Erwägungen im Sinne der administrativen Effizienz vorgehen.

Einen letzten Willkürfaktor bei der Sprachkompetenzmessung gilt es zu benennen: Die Zeit, die seit dem Ablegen und Bestehen einer Prüfung verstrichen ist: Hier legen einzelne Testanbieter bestimmte Zeiträume für die „Gültigkeit“ ihrer Zertifikate fest, andere nicht. Für die relevanten anerkannten DaF/DaZ-Prüfungsanbieter gilt zwar, dass ihre Zertifikate im Unterschied etwa zum TOEFL für den englischsprachigen Hochschulzugang unbegrenzt gültig sind,⁷ diese Festlegung wird jedoch durch die deutschen Botschaften oder einzelnen Hochschulen ausgehebelt, wenn sie verlangen, dass die Sprachnachweise nicht älter als ein oder zwei Jahre sein dürfen.

⁷ telc- und TestDaF-Zertifikate sind unbegrenzt gültig: https://www.telc.net/fileadmin/user_upload/Informationsmaterial/leitfaden_fuer_schulen.pdf, S. 9; <https://www.testdaf.de/de/teilnehmende/mein-testdaf/faq/faq-ergebnisse-und-zertifikat/>.

Das Visumshandbuch (Auswärtiges Amt 2022) verlangt, dass geforderte Sprachnachweise nicht älter als ein Jahr sind. Worauf sich diese Zahlen gründen, ist unklar, es gibt keine allgemeingültigen wissenschaftlichen Untersuchungen, in welchem Tempo Sprachkompetenz verloren geht, und selbst wenn, wie der Testanbieter für den TOEFL, der Educational Testing Service (ETS), behauptet, bleibt jeder konkrete Zeitpunkt willkürlich. Der Sprachkompetenzverlust dürfte auf jeden Fall ein individuell sehr unterschiedlicher und komplexer Prozess sein.

Relevant ist vor allem, wie schnell die einmal Getesteten wieder das zertifizierte Niveau erreichen bzw. ob durch Lebensumstände des Zertifikatsinhabenden begründet die Sprachkompetenz aufrechterhalten wurde oder vielleicht inzwischen sogar höher als zum Testzeitpunkt ist. Da jeder summative Test also nur eine Momentaufnahme bescheinigen kann, wird er im Hinblick auf die Zwecke des Nachweises wie sprachliche Studier- oder Arbeitsfähigkeit de facto aus Sicht des Nachweisfordernden doch zu einem formativen Potenzialtest.

Natürlich kann man sich auch fragen, ob z.B. Visa zum Zwecke des Studiums nicht generell (wieder) ohne bereits existierende Sprachnachweise (in den meisten außereuropäischen Ländern liegen die verlangten Sprachnachweise zwischen A1 und B1) erteilt werden sollten, wenn die Studienabsicht ansonsten plausibel ist, so wie das früher der Fall war.

Der Grund, dass das nicht mehr so ist, liegt in einer zunehmenden Abschottungspolitik der BRD gegenüber Zuwanderung aus nicht-europäischen Ländern bzw. westlichen Industriestaaten. Zugegebenermaßen ist es auch nicht von der Hand zu weisen, dass Visa zum Zwecke des Studiums auch oft ohne ernsthafte Studienabsicht beantragt werden bzw. wenn man dann – nicht zuletzt aus sprachlichen Gründen – keinen Studienabschluss schafft, diese Studierenden daraufhin unter hohem Druck stehen, nicht als Versager:in zurückzukehren, so dass sie in die Illegalität wechseln oder Asyl beantragen. Es gibt gute Gründe für die Koppelung der Visaregelungen an Sprachnachweise, denn die Plausibilität, dass so Kandidat:innen ausgefiltert wurden, die Probleme mit dem Deutscherwerb und folglich mit einem Studium auf Deutsch hätten, ist nicht von der Hand zu weisen.

Festzuhalten bleibt aber zum einen, dass so auch begabte und motivierte Studienbewerber:innen eliminiert werden, die vielleicht nur das Pech hatten, kein Geld für einen Deutschkurs zu haben bzw. kein Angebot in ihrer Nähe zu finden. Wenigstens der letzte Punkt dürfte durch Online-Angebote eine zunehmend geringere Rolle spielen. Zum anderen ist es offensichtlich, dass hier wie auch in anderen Fällen Sprachnachweise als Instrument zur Steuerung der Migrationsströme zweckentfremdet werden, wie mehrfach erlebt, wenn in einzelnen Ländern bei steigenden Bewerberzahlen, das Niveau des Deutschnachweises hochgesetzt wurde.

7 Zwischenstand

Fassen wir zusammen: Die existierenden, offiziell und das heißt von den Behörden anerkannten Sprachprüfungen sind in manchen Hinsichten eher individuell willkürliche Instrumente zur Steuerung von Migrationsströmen als unfehlbare Diagnosen, wer sprachlich erfolgreich in Deutschland in Studium und Beruf bestehen könnte. Sie verbauen im Einzelfall Personen den Weg nach Deutschland, die sich hier sprachlich erfolgreich integrieren würden, und gewähren anderen Zugänge, die aufgrund ihrer Deutschkompetenzen und persönlichen Umstände scheitern werden. Dass standardisierte Prüfungen nicht unfehlbar sind, wissen auch die Prüfungsanbieter.

Schwerer als Mängel der Testkonstrukte, qualitativ schlecht gemachte Prüfungssätze, problematische Testcenter, die Prüfungen nicht ordnungsgemäß durchführen oder Schwächen in der Bekämpfung von Täuschungsversuchen (Testsicherheit), die sich nie ganz vermeiden lassen, wiegen die prinzipiellen Einwände. Standardisierte Tests müssen große Stückzahlen erreichen, folglich verallgemeinern, so dass sie auf möglichst viele Situationen irgendwie passen, aber auf keine so ganz. Das haben wir oben schon am Beispiel des Fertigkeitenmixes von Prüfungen gezeigt, aber es betrifft auch die Prüfungsinhalte. Die sprachlichen Hochschulzugangsprüfungen etwa müssen auf alle Fächer bzw. Studieninteressen passen, die Texte und Aufgaben berühren zwar deshalb meist mehrere Fachgebiete, benachteiligen somit aber alle, die z.B. ein eng technisches, aber wenig gesellschaftliches Verständnis haben (und umgekehrt) bzw. privilegieren diejenigen, deren Interessen breit gestreut sind. Viel besser geeignet und fairer wären fachbezogene Sprachprüfungen, was aus offensichtlichen Gründen angesichts des komplexen Prozesses, an dessen Ende ein Studium in Deutschland aufgenommen wird, nicht praktikabel wäre. So konzentrieren sich die standardisierten akademischen Sprachtest besonders gerne auf das allgemeine Studiervokabular und allgemeine Szenarien aus dem Hochschulleben.

Der Praktikabilität und das heißt auch dem im jeweiligen Land vorhandenen Netzwerk an Sprachkursen, Prüfungsterminen und Prüfenden ist es auch geschuldet, dass im Zuge der zunehmenden Fachkräfterekrutierung im Ausland weiterhin fast ausschließlich die allgemeinsprachlichen Prüfungen, deren Lesetexte und Schreibaufgaben immer noch merklich von einem europäisch-bildungsbürgerlichen Zielpublikum ausgehen, abgenommen werden, die wenig auf die neue Zielgruppe der Arbeitsmigrant:innen und deren Bedürfnisse zugeschnitten sind. Das führt zu so unsinnigen Effekten, dass in derartigen Programmen zwar die Berufssprache vermittelt werden soll, was für die Teilnehmenden immer besonders motivierend ist, die Migrationswilligen aber faktisch allein auf die visumsrelevanten allgemeinsprachlichen Prüfungen hin gedrillt werden müssen, ohne deren Bestehen sie nicht ausreisen dürfen.

Reliabilität und Validität wie auch die Testsicherheit standardisierter Prüfungen sicherzustellen, bei denen es für die Prüflinge um viel geht (die sogenannten High-Stakes-Prüfungen), ist ein permanenter und wichtiger Prozess. Hier kann man Details

kritisieren, arbeitet vielleicht der eine Testanbieter etwas sorgfältiger als der andere und natürlich entwickelt sich die Testtheorie auch weiter. Mir ging es bisher vor allem darum, den blinden Glauben an die Messgenauigkeit dieser Prüfungen für den Spracherwerbsprozess und die kommunikativen Erfolgsprognosen in Frage zu stellen. Diese Erkenntnisse sind nicht neu, scheinen aber durch technokratischen Steuerungsdruck zunehmend ins Hintertreffen zu geraten.

8 Sprachnachweise zur Integrationssteuerung

Auch in der Inlandsperspektive oder vereinfacht gesagt in der DaZ-Perspektive spielen Sprachtests, die Kompetenzen im Sinne der Stufen des Referenzrahmens nachweisen, eine dominante Rolle, da sie als Maß zur Beurteilung der Integrationschancen Erwachsener häufig über die Gewährung staatlicher Förderungen und Sozialleistung, Staatsbürgerschaft, Abschiebung, Berufsausübung u.a.m. entscheiden. Die Einführung des „Deutstests für Zuwanderer“ (DTZ) als einheitliche Abschlussprüfung für die Integrationskurse (A1–B1) im Jahr 2009 und des „Deutstests für den Beruf“ (DTB) als Abschluss der Berufssprachkurse (A2–C1) im Jahr 2022 zeigen nicht nur den gesteigerten Glauben an die Objektivität von Sprachtests, sondern auch wie wichtig dem Staat die Kontrolle in diesem Bereich geworden ist, und das heißt auch, wie sehr er an die höhere Validität „seiner“ Tests glaubt. Warum sollte er sich sonst nicht mit den bereits existierenden standardisierten Tests zufriedengeben?

Zu den grundsätzlichen Einwänden gegen den allzu naiven Glauben an die uneingeschränkte Validität von Sprachtests als zielgruppengerechtes Instrument zur Messung kommunikativer Kompetenzen und erst recht von Potenzialen für bestimmte Zwecke, das oben vor allem bezogen auf den Bereich der Sprachprüfungen für den Hochschulzugang ausgeführt wurde, kommen bei der Integration von Migrant:innen, die bereits in Deutschland leben, weitere Aspekte hinzu.

Eigentlich sollte man denken, dass das Lernen einer neuen Sprache in der zielsprachlichen Umgebung leichter fällt und auch jede:r Erwachsene jedes Sprachniveau erreichen kann. Dem steht das bekannte Phänomen der Fossilierung entgegen. Damit ist gemeint, dass nach einer zunächst erreichten, und sei es minimalen Sprachkompetenz, keine Lernfortschritte mehr gemacht werden. Die Gründe sind vielfältig und die Erklärung des Phänomens ist weitgehend spekulativ. Eine Rolle spielen wohl Faktoren wie Alter, Motivation, allgemeines Sprachbewusstsein und allgemeine sprachliche Kompetenz in der Erstsprache, dort vor allem auch Beherrschung der Schrift- und Bildungssprache, andere Fremdsprachen, das Maß der exklusiven Einbindung in der eigenen erstsprachlichen Community, die verschiedenen, auch non-verbalen „Umgehungsstrategien“, um sprachlichen Defizite zu kompensieren u.a.m. Problematisch ist das vor allem dann, wenn Fossilierungen auf einem niedrigen Niveau eintreten. Dieses Phänomen ist in den Integrations- und Berufssprachkursen, aber auch im Alltag häufig zu beobachten, übrigens auch bei

hochqualifizierten Expatriates oder Rentner:innen, die auch nach langen Jahren in einem anderen Land, in dem sie schließlich sogar ihren Lebensabend verbringen, nur einige Chunks der dort geltende Verkehrssprache aktiv beherrschen.

Die veröffentlichten Daten zu den Integrationskursen in Deutschland zeigen auch bei mehrfacher Testteilnahme für das Jahr 2021 eine Bestehensquote von 60,2% beim B1-Test D'TZ (vgl. BAMF 2022). Allerdings bezieht sich diese Prozentzahl nur auf die Teilnehmenden, die sich überhaupt zur Prüfung anmelden und berücksichtigt auch nicht diejenigen, die vorher abbrechen. Für die A2- und B1-Berufssprachkursprüfungen, d.h. für Migrant:innen, die vermutlich bereits an mehr als 1.000 UE Deutschunterricht teilgenommen haben, wurden seit 2017 noch keine Bestehensquoten veröffentlicht.

Das heißt: Trotz staatlich intensiv und großzügig geförderten Sprachunterricht, der bei Nicht-Bestehen der Prüfungen weit über 1.000 UE umfassen kann, kommen vermutlich weit mehr als die dokumentierten 39,8% der Migrant:innen in staatlich geförderten Sprachkursen bei standardisierten Tests nicht über das Niveau A2 hinaus, 12,2% der Teilnehmende erreichen nicht einmal A2, von denen allerdings der Großteil aus den Alphabetisierungskursen stammt.

Es ist fragwürdig, diese Erwachsenen immer wieder in Prüfungen zu schicken, die sie nicht bestehen können, weil sie keine Fortschritte mehr im Sinne des Testkonstrukts, speziell in den schriftbasierten Fertigkeiten mehr machen (können). Die erfolgreiche sprachliche Integration nach Definition des Gemeinsamen Europäischen Referenzrahmens für Sprachen, d.h. das Erreichen der Niveaus B2/C1 bzw. erstsprachlicher Kompetenz ist in vielen Fällen eine Frage, die mehr als eine Generation benötigt (vgl. Jung 2017: 52f., 61). Trotzdem kommen die Betroffenen möglicherweise in ihrem Lebensbereich im Alltag gut klar und üben auch ihren Beruf in einem deutschsprachigen Kontext erfolgreich aus.

Vielleicht sollte man die Perspektive einmal umdrehen und versuchen die offensichtlich dennoch vorhandenen sprachlichen Fähigkeiten in realen Situationen zu erfassen, um so zu einem individuellen Kommunikations- und Kompetenzprofil zu gelangen, statt durch die bildungssprachliche Brille betrachtet immer wieder nur „Defizitbescheide“ in Testform zu erstellen. Das wäre vermutlich ein wesentlich sinnvollerer Einsatz von Lebenszeit und Personalressourcen als weitere Standardkurse mit Standardprüfungen und zudem ein guter Ausgangspunkt für ein berufsbegleitendes Sprachcoaching. Dadurch ließen sich Fossilierungen aufgrund der situativen Einbettung leichter aufbrechen, um so einen echten Kompetenzzuwachs über A2 hinaus vor allem in den mündlichen Fähigkeiten zu erreichen, was eine Vierfertigkeitenprüfung immer noch als „nicht-bestanden“ werten müsste.

9 Fazit und Ausblick

Dieser Beitrag ist kein Plädoyer gegen summative, standardisierte Deutschprüfungen. Sie sind in vielen Fällen eine notwendige und hilfreiche Sprachstandsdiagnose und oft die vergleichsweise fairste Form des Sprachnachweises. Gute Tests erfordern eine hohe theoretische, konzeptionelle und handwerkliche Kompetenz der Tester:innen und der wissenschaftliche Fortschritt wird sie weiter verbessern. Im Zuge der allgegenwärtigen Implementierung von Deutschtests mit weitreichenden Folgen für die Prüflinge geht es mir darum, an die nicht-hintergehbaren Schwächen standardisierter Sprachtests zu erinnern, ihre Objektivität zu relativieren und vor ihrer allzu unkritischen Verwendung als objektives Steuerungsinstrument zu warnen. Standardisierte Sprachprüfungen sind häufig unfair, werden vielen ein bisschen, aber wenigen ganz gerecht. Speziell im Bereich der Fachkräfteanwerbung und der Geflüchtetenintegration ist erstens in bestimmten Fällen über eine Reduktion des Prüfungsmarathons und den Ersatz von Prüfungen durch die Erhebung von Kompetenzprofilen und zweitens über den verstärkten Einsatz von Szenarioprüfungen nachzudenken, wie sie Mediziner:innen, und Apotheker:innen und demnächst auch Pflegekräfte vor den zuständigen Gremien ablegen müssen. Was diese durch die Zufälligkeiten des Simulationsverlaufs mit Schauspielpatient:innen bzw. der Jurys an Vergleichbarkeit bzw. Reliabilität verlieren, gewinnen sie an Validität hinsichtlich arbeitsplatzspezifischer Kommunikationskompetenz. Dieses Modell ließe sich auch auf viele Mangelberufe übertragen und könnte mit einer Jury von Arbeitgeber:innen der betreffenden Branche und Sprachexpert:innen sogar online, zumindest alternativ zu den standardisierten Sprachtests durchgeführt werden.

So oder so: Die vier Jahrzehnte alte Bonmot von Spolsky (1981, zit.n. ALTE 2012: 22), dass vor allem wichtige Tests wie Medikamente mit dem Hinweis „Mit Vorsicht einzusetzen“ versehen sein sollten, ist aktueller denn je – ein Diktum, dem sich vermutlich auch Uwe Koreik anschließen würde, obwohl man davon ausgehen kann, dass es bei den von ihm verantworteten Prüfungen dieser Warnung nicht bedarf.

Literatur

ALTE (2012): *Handbuch zur Entwicklung und Durchführung von Sprachtests. Zur Verwendung mit dem GER*. Erstellt von ALTE im Auftrag des Europarats/Abteilung für Sprachenpolitik. Frankfurt a.M.: telc.

Auswärtiges Amt (2022): *Visumshandbuch*: Stand: 22. Ergänzungslieferung März 2022. (<https://www.auswaertiges-amt.de/blob/207816/101441d43c4bbac8da90ad8bb00903be/visumhandbuch-data.pdf>. Letzter Zugriff: 15.09.2022).

- BAMF (2022): Bericht zur Integrationskursgeschäftsstatistik für das Jahr 2021. (https://www.bamf.de/SharedDocs/Anlagen/DE/Statistik/Integrationskurszahlen/Bundesweit/2021-integrationskursgeschaeftsstatistik-gesamt_bund.pdf?__blob=publicationFile&v=2. Letzter Zugriff:12.09.2022).
- Biggs, John; Tang, Catherine (2011): *Teaching for Quality Learning at University*. Maidenhead: Open University Press.
- British Council, UKALTA, EALTA, ALTE (2022): *Aligning Language Education with the CEFR: A Handbook*. Published jointly by the European Association for Language Testing and Assessment (EALTA), the UK Association for Language Testing and Assessment (UKALTA), the British Council, and the Association for Language Testers in Europe (ALTE). (https://www.britishcouncil.org/sites/default/files/cefr_alignment_handbook_layout.pdf. Letzter Zugriff:12.09.2022).
- Casper-Hehne, Hiltraud; Koreik, Uwe (Hrsg.) (2004): *DSH und TestDaF als hochschulbezogene Prüfungssysteme für Deutsch als Fremdsprache – Standortbestimmungen und Entwicklungslinien*. Hohengehren: Schneider.
- Council of Europe (2009): *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A Manual*. (<https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d>. Letzter Zugriff: 15.09.2022).
- DGFF (2018): *Reform, Remake, Retusche? – Diskussionspapier der Deutschen Gesellschaft für Fremdsprachenforschung zum Companion to the CEFR* von Olaf Bärenfänger, Claudia Harsch, Bernd Tesch und Karin Vogt. (<https://www.dgff.de/assets/Uploads/DGFF-Stellungnahme-zum-Companion-to-the-CEFR-2018.pdf>. Letzter Zugriff:11.09.2022).
- Koreik, Uwe (Hrsg.) (2005): *DSH und TestDaF eine Vergleichsstudie*. Baltmannsweiler: Schneider Verlag Hohengehren.
- Jung, Matthias (2017): Sprachliche Förderung von Flüchtlingen und Migranten. Kontinuitäten und Diskontinuitäten seit den 70er Jahren. In: Barz, Heiner (Hrsg.): *Flüchtlinge willkommen und dann? Die Flüchtlingskrise als Herausforderung für Gesellschaft und Bildung*. Düsseldorf: University Press, 41–63.
- Zeh, Juli (2012): *Nullzeit*. Frankfurt a.M.: Schöffling.