

Die Vielfalt des aktuellen Testgeschehens – Vorwort

<https://doi.org/10.1515/infodaf-2022-0063>

Wie es auch schon einleitend im vorangegangenen Themenheft „Testen und Prüfen“ (Heft 4, 2022) hieß, hat es in den letzten zwei Jahrzehnten in allen Bereichen „des gesellschaftlichen, politischen und wirtschaftlichen Lebens tiefgreifende Veränderungen gegeben, die sich auch auf den Bereich des Testens und Prüfens von Sprachkenntnissen auswirkten“ (Kniffka/Koreik 2022: 285) und deswegen auch immer wieder der Revision bedürfen. Viele der im Vorwort der letzten Ausgabe thematisierten Aspekte wie Auswirkungen des Fachkräfteeinwanderungsgesetzes, der Bedarf für die Überprüfung berufsbezogener sprachlicher Kompetenzen im Rahmen der Integration Geflüchteter auf den deutschen Arbeitsmarkt oder die schulische und damit letztlich auch gesellschaftliche Integration von Kindern mit Deutsch als Zweitsprache und anderes ließen sich auch hier anführen. Es soll aber lediglich ein weiterer Aspekt hervorgehoben werden. Im renommierten Standardwerk „Deutsch als Fremdsprache: Ein internationales Handbuch“ (Helbig et al. 2001) hieß es in dem für unsere Thematik relevanten Artikel von Perlmann-Balme zu „Formen und Funktionen von Leistungsmessung und -kontrolle“ folgendermaßen:

„Das in den Vereinigten Staaten vorherrschende Misstrauen gegenüber intuitiv korrigierten Aufsätzen sowie die Skepsis hinsichtlich der Objektivität von Lehrkräften bei der Beurteilung der Leistungsfähigkeit ihrer eigenen Schüler wird im deutschsprachigen Raum nicht geteilt. In Deutschland herrscht sogar ein Misstrauen gegen zentral verordnete Prüfungen. [...] Deutsche Autoren werfen den Angelsachsen gern eine Überbewertung der vom Behaviorismus geprägten Formen der Leistungsmessung vor, und wenden dagegen ein, dass sie die ganzheitliche Lernerpersönlichkeit außer Acht lasse“ (Perlmann-Balme 2001: 995).

Und dann kam der PISA-Schock, der in Deutschland eine bildungspolitische Debatte auslöste und auch deutliche Auswirkungen auf die Einstellung zu Tests, Testgestaltung und Gütekriterien der Testbewertung hatte. In dem entsprechenden Artikel der gleichen Autorin in der aktualisierten Neuauflage des Handbuchs, nun unter dem Titel „Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch“ erschienen (Krumm et al. 2010), hieß es in dem entsprechenden Artikel „Testen und Prüfen von Sprachkenntnissen“ dann an weitgehend identische Textpassagen angefügt etwas vorsichtiger formuliert:

„Doch diese Ablehnung weicht allmählich der Erkenntnis, dass gemeinsame, objektive Maßstäbe notwendig sind, will man Leistungen vergleichen“ (Perlmann-Balme 2010: 1273).

Und diese Erkenntnis hat sich nun, weitere zwölf Jahre später, in den beiden Themenheften zum Testen und Prüfen deutlich niedergeschlagen, in denen sehr offensichtlich wird, dass die Fachdiskussion sich hinsichtlich der Testtheorie und damit der Testerstellung und Testbewertung (auch) in unserem Fach auf einem erheblich höheren Niveau bewegt, als das vor zwanzig Jahren denkbar war. Dies gilt, auch wenn zunehmend deutlich wird, dass eine immer kleinschrittigere Orientierung auf Kompetenzen vielleicht tatsächlich das Problem aufwerfen kann, dass eine „ganzheitliche Lernerpersönlichkeit außer Acht“ gelassen wird.

Auf jeden Fall ist im Fach Deutsch als Fremd- und Zweitsprache im Bereich „Testen und Prüfen“ eine deutliche Qualitätsentwicklung mit Hinblick auf Professionalisierung zu verzeichnen, was seinen Ausdruck im letzten und in diesem Themenheft von Info DaF findet. In der wissenschaftlichen Diskussion finden sich inzwischen eine deutlich gehäufte Zahl an Untersuchungen, eine Pluralität an thematischen Herangehensweisen und eine Vielfalt an sehr diversen Lernergruppen, für deren Bedarfe auf testtheoretischer und testpragmatischer Ebene Untersuchungen und Überlegungen angestellt werden, wie sie vor zwanzig Jahren nicht vorstellbar gewesen wären. Das Spektrum reicht von Tests im Kita-Bereich über Tests im schulischen Kontext (Stichwort: DaZ), über berufs- oder fachsprachenspezifische Tests bis zu den traditionell eher etablierten Diskussionen über Tests zur Hochschulzulassung; als innovativ erweisen sich Tests, die die Leistung von Menschen erfassen sollen, die nicht oder nur bedingt alphabetisiert sind.

Bei der Zuwanderung in die deutschsprachigen Länder gibt es nicht wenige Menschen, die entweder mit dem lateinischen Schriftsystem nicht ausreichend vertraut oder sogar in der eigenen Herkunftssprache nicht alphabetisiert sind. Um diese Gruppe zuverlässig identifizieren und zielgerichtete Förderentscheidungen einleiten zu können, hat das Amt für Wirtschaft und Arbeit des Kantons Zürich einen Subtest zur Deutscheinschätzung für fremdsprachige Stellensuchende entwickeln lassen. Dieser optionale Subtest ist ein diagnostisches Instrument, das sich an die Beschreibung unterschiedlicher Niveaustufen von Alphakompetenz des LEA-Projekts anlehnt und im Beitrag von **Olaf Bärenfänger** vorgestellt und analysiert wird. Im Vordergrund steht das Vorgehen bei der Erstellung eines diagnostischen Instruments zur Erfassung von unzureichender Schriftkompetenz. Dabei spielen auch einige Besonderheiten von berufsbezogener Sprachverwendung eine Rolle. Das Testkonstrukt wird mit Blick auf die konkreten Aufgabenstellungen und Bewertungskriterien sowie hinsichtlich der Validierung des Prüfungsteils „Alpha“ bewertet und abschließend als diagnostisches Instrument für abgestimmte Fördermaßnahmen eingeschätzt.

Deutsch im Beruf ist mit zunehmender Arbeitsmigration und Arbeitskräftemobilität ein zentrales Thema im DaF/DaZ-Bereich geworden. Und damit spielen auch Tests für berufsbezogene Sprachkenntnisse eine immer größere Rolle. Ein Test zum Nachweis berufsbezogener Sprachkenntnisse ist in diesem Zusammenhang der Goethe-Test PRO: Deutsch für den Beruf (GTP). **Aron Fink, Andreas Frey und Katharina Klein** gehen in ihrem Artikel zunächst der berufsorientierten Ausrichtung des Tests nach und widmen sich eingehend dem Testformat des GTP und damit vor allem auch der Beschreibung seiner psychometrischen Güte. Ihre Analyse und ihre Ergebnisse werden abschließend mit Blick auf die Relevanz von berufsbezogenen Tests für Deutsch als Fremdsprache einer Diskussion unterzogen.

Alice Friedland und Milica Sabo charakterisieren die DSH (Deutsche Sprachprüfung für den Hochschulzugang) als eine Prüfung, „die nicht grundsätzlich standardisiert ist und dementsprechend dezentral erstellt wird bzw. werden kann“, verdeutlichen in ihrem Beitrag aber zugleich, welche Fortschritte bei der Qualitätssicherung der DSH in den letzten Jahren (mit dem Start 2005) vor allem durch die Initiative des FaDaF (Fachverband Deutsch als Fremdsprache) gemacht wurden und wie dabei jeweils vor Ort vorgegangen werden kann.

Am Beispiel des DSH-Standortes Friedrich-Schiller-Universität Jena wird herausgestellt, wie wichtig die professionelle Erstellung eines Tests und der Bewertungskriterien ist und welche Rolle die sogenannte Ober- und Untergruppenanalyse bzw. die vereinfachte Itemanalyse für die Evaluation des Tests wie auch hinsichtlich der Gütekriterien Schwierigkeit und Trennschärfe spielen. Dabei werden vier DSH-Prüfungssätze exemplarisch zugrunde gelegt, an denen für die Prüfungsteile der rezeptiven Fertigkeiten Leseverstehen und Hörverstehen die Relevanz aufgezeigt wird. Damit wird akribisch ein Vorschlag zur Testevaluation gemacht, welcher auch an anderen DSH-Standorten umgesetzt werden kann.

Mirka Mainzer-Murrenhoff und Anastasia Drackert beschäftigen sich mit dem inzwischen seit Jahrzehnten bekannten C-Test-Format und stellen es einleitend zunächst kurz dar, beleuchten es hinsichtlich seiner Entstehungsgeschichte, verweisen zu Recht auf den zurzeit noch aktuellsten Überblick bei Grotjahn (2019) und erörtern verschiedene Einsatzmöglichkeiten dieses Testformats. Ihre Feststellung, dass der Einsatz dieses Testverfahrens für Anliegen im Bereich „Sprachbildung und DaZ“ hingegen vergleichsweise neu sei, führt zu einer Diskussion über die Einsatzmöglichkeiten in diesem Feld. Um dieser Frage nachzugehen, werden zunächst die bisherigen (vorrangigen) Anwendungsbezüge des C-Tests erörtert, die bisherigen Adressatenkreise in den Blick genommen und das Testkonstrukt

selbst sowie die Konzeption des C-Tests zum Thema gemacht. Abschließend werden Desiderate unter dem Aspekt von Untersuchungen zum Testkonstrukt und zu Zwecken des Einsatzes des C-Tests im Arbeitsfeld „Sprachbildung und DaZ“ diskutiert.

Eine ganz andere Zielgruppe haben **Pascale Schaller** und **Julia Winkes** im Blick, wenn sie den Ansatz des Curriculum-basierten Messens (CBM) in den Fokus rücken und diesen im Kontext der Lernfortschrittsdiagnostik für Kinder mit Deutsch als Zweitsprache behandeln. Behandelt wird konkret die Zielgruppe von Kindern mit Deutsch als Muttersprache und Deutsch als Zweitsprache der 3. bis 6. Klassenstufe anhand einer Stichprobenuntersuchung. Im Vordergrund der Betrachtungen stehen die Reliabilität und die Validität des Verfahrens. Entscheidende Fragestellungen sind dabei, inwieweit sich die Ergebnisse auf Kinder mit Deutsch als Zweitsprache übertragen lassen und ob tatsächlich eine einzelne dreiminütige Schreibprobe ausreichend reliabel ist. Um diese Fragen beantworten zu können, muss allerdings zunächst CBM Schreiben vorgestellt und die Funktion einer Lernverlaufsdagnostik geklärt werden. Es gibt interessante Ergebnisse, aber die Limitationen der Studie werden auch klar benannt und damit die Basis für weitere Studien gelegt.

Almut Schön rückt Fachsprachenprüfungen für Ärzte, Apotheker und Pflegekräfte in den Vordergrund, also Prüfungen, die es in dieser Form, hervorgerufen durch den Fachkräftemangel und den damit verbundenen Druck, in diesen Bereichen verstärkt Zuwanderung zu ermöglichen, erst seit einigen Jahren gibt. Sie wirft dabei die Frage auf, inwieweit „diese Prüfungen die Hauptgütekriterien von Tests erfüllen und welche Forschungsaufgaben sich hier für die angewandte DaF-Forschung ergeben.“ Sie kommt in ihrer Untersuchung, in der sie die Gütekriterien für Tests zum Maßstab macht, zu einem komplexen und differenzierten Ergebnis, hebt allerdings hervor, dass die Erhebung und statistische Auswertung größerer Mengen an Prüfungen nötig wäre, um damit qualifiziertere Aussagen zur Durchführungsobjektivität und vor allem auch zur Reliabilität treffen zu können.

Lisa Teufele beschäftigt sich wie bereits im Titel ihres Artikels angekündigt mit der Validität berufsbezogener Sprachtests. Ausgehend vom englischen Begriff des *specific purpose language testing* und unter der Maxime, dass „der *specific purpose*, der sich auf das Testkonstrukt eines berufsbezogenen Sprachtests bezieht, klar definiert sein muss“, richtet sie die Forderung an Sprachtestentwickler*innen, dass die von ihnen erstellten Tests zweck- und kontextgebundene berufs-/arbeitsplatzrelevante Sprachkenntnisse und -kompetenzen zu erfassen hätten. Das Fazit ihrer Analyse ist dann eher ernüchternd, wenn sie eine häufig

fehlende Überprüfung der Konstruktvalidität bei berufsbezogenen Sprachtests konstatiert und diesen methodische Defizite attestiert.

Bielefeld und Freiburg im Oktober 2022

Gabriele Kniffka

Uwe Koreik

Literaturverzeichnis

- Grotjahn, Rüdiger (2019): „C-Tests“. In: Jeuk, Stefan; Settinieri, Julia (Hrsg.): *Sprachdiagnostik Deutsch als Zweitsprache: ein Handbuch*. Berlin: De Gruyter, 579–603.
- Helbig, Gerhard; Götze, Lutz; Henrici, Gert; Krumm, Hans-Jürgen (Hrsg.) (2001): *Deutsch als Fremdsprache: Ein internationales Handbuch*. Berlin: De Gruyter.
- Kniffka, Gabriele; Koreik, Uwe (2022): „Vorwort“. In: *Informationen Deutsch als Fremdsprache* 49 (4): 285–288 (Themenheft „Testen und Prüfen“).
- Krumm, Hans-Jürgen; Fandrych, Christian; Hufeisen, Britta; Riemer, Claudia (Hrsg.) (2010): *Deutsch als Fremd- und Zweitsprache: Ein internationales Handbuch*. Berlin: De Gruyter.
- Perlmann-Balme, Michaela (2001): „Formen und Funktionen von Leistungsmessung und -kontrolle“. In: Helbig, Gerhard; Götze, Lutz; Henrici, Gert; Krumm, Hans-Jürgen (Hrsg.), 994–1006.
- Perlmann-Balme, Michaela (2010): „Testen und Prüfen von Sprachkenntnissen“. In: Krumm, Hans-Jürgen; Fandrych, Christian; Hufeisen, Britta; Riemer, Claudia (Hrsg.), 1272–1288.